



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 117402

TO: Manjunath N Rao
Location: REM/3B81/3C70
Art Unit: 1652
Monday, March 29, 2004

Case Serial Number: 10/026140

From: Noble Jarrell
Location: Biotech-Chem Library
Rem 1B71
Phone: 272-2556

Noble.jarrell@uspto.gov

Search Notes

This Page Blank (uspto)

From: Rao, Manjunath N.
Sent: Monday, March 22, 2004 11:41 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 10/026,140

From: Manjunath N. Rao
Art Unit 1652, Room 3B81
Mail Box in Room 3C70
Phone: 272-0939

Date: 3-22-04

Please search the following as soon as possible for application with serial number

10/026,140

1. SEQ ID NO: 1 and 3 against all commercial nucleic acid databases, issued patents/published applications database and pending application database. Please provide a print of all results
2. SEQ ID NO: 2, against all commercial amino acid databases, issued patents/published applications database and pending application database. Please provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao
Art Unit 1652, Room 3B81,
Mail Box in Room 3C70,
Remsen Building, USPTO

Searcher: Janrell
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 3/29/04
Searcher Prep/Review: 3
Clerical: _____
Online time: 60

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: 1
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: Compu gen
WWW/Internet: _____
Other (specify): _____

RECEIVED
MAR 22 2004
STIC

This Page Blank (uspto)

400, Dulany St.
Alexandria, VA.
Phone: 571-272-0939

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

This Page Blank (uspto)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.





This Page Blank (uspto)


```
DB 121 TYHMDLPQALHNRVGGWMLNVEEYQDPERFARALCEGRGVDVQWITINXWIAITGY 180
OY 181 ATGSNAPGSSSTINKSTTECNTATEPRLAKQAQIMSHARAVAVYSDPSPQKQIGISLN 240
DB 181 ATGSNAPGSSSTINKSTTECNTATEPRLAKQAQIMSHARAVAVYSDPSPQKQIGISLN 240
OY 241 GDYEPWDSNEPRDKEAERMEFHIIGFANPIFLKKOYPSMKKQOLGERLPLALPAPA 300
DB 241 GDYEPWDSNEPRDKEAERMEFHIIGFANPIFLKKOYPSMKKQOLGERLPLALPAPA 300
OY 301 IINAGETDFYGMNYYTSQFARHLDGPVPEPTYLGAIHHEQNKDGS PVGESGLAMLRSC 360
DB 301 IINAGETDFYGMNYYTSQFARHLDGPVPEPTYLGAIHHEQNKDGS PVGESGLAMLRSC 360
OY 361 PMFMRKHARVYGLGKPIYITENGCPCGGEENMTCEAVNDPFRIRYFDSHLSISKA 420
DB 361 PMFMRKHARVYGLGKPIYITENGCPCGGEENMTCEAVNDPFRIRYFDSHLSISKA 420
OY 421 TODGVVVGKGFAMALLDNLNLEWSDGYGPRFGVTFTDYTLAKTPPKKSALVLDKMFARQV 480
DB 421 TODGVVVGKGFAMALLDNLNLEWSDGYGPRFGVTFTDYTLAKTPPKKSALVLDKMFARQV 480
OY 481 KVAA 484
DB 481 KVAA 484

RESULT 2
ID 093784 PRELIMINARY; PRT; 476 AA.
AC 093784;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Beta-glucosidase.
GN BGL4.
OS Humicola grisea var. thermoidae.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=5528;
RN [1]
RP SEQUENCE FROM N.A.
RA Takashima S., Nakamura A., Hidaka M., Masaki H., Uozumi T.,
RT "Molecular cloning and expression of the novel fungal beta-glucosidase
RL genes from Humicola grisea and Trichoderma reesei."
J. Biochem. 0:0-0(1999).
DR EMBL; AB003109; BAA74958.1; -.
DR HSSP; P22073; ITR1.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
SQ SEQUENCE 476 AA; 54062 MW; DA17ADF5CB9935 CRC64;

Query Match 49.3%; Score 1303.5; DB 3; Length 476;
Best Local Similarity 51.7%; Pred. No. 2.5e-96;
Matches 244; Conservative 79; Mismatches 144; Indels 5; Gaps 5;
```

```
OY 185 NAPSSTINKSTTECNTATEPRLAKQAQIMSHARAVAVYSDPSPQKQIGISLNDYY 244
DB 180 PAPHTSDRSKSPVGSARSPWIVGHNIIIAHRAVAVAYDEPKPGGEGITLNGDAT 239
OY 245 EPMDSNEPRDKEAERMEFHIIGFANPIFLKKOYPSMKKQOLGERLPLALPAPA 304
DB 240 LPMDEPDAIEACDRKIEFAISMFADPIYFGK-YFDSMKQAGDRLEPTEPPEVALVK- 297
OY 305 GETDFYGMNYYTSQFARHLDGPVPEPTYLGAIHHEQNKDGS PVGESGLAMLRSCPDMP 364
DB 298 GSNDFYGMNHYTANYIKHKTGVPEDEDFLGNETLFPNKYGDICGETQSFMLRPHAGP 357
OY 365 RKHLARVYGLGKPIYITENGCPCGGEENMTCEAVNDPFRIRYFDSHLSISKAITD 423
DB 358 RDLNMTSKKRYGPKIYVENGSLKGEENMPLEQVLEDFRYKYNNDVYRAAAAVAAD 417
OY 424 GVVVVGKGFAMALLDNLNLEWSDGYGPRFGVTFTDYTT-LKTPPKKSALVLDKMF 474
DB 418 GNVVGYIAMSLLDNFEMAGETFRGVTVYVYANDOKYPPKKSALSLPLF 469

RESULT 3
ID 08X214 PRELIMINARY; PRT; 489 AA.
AC 08X214;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Beta-glucosidase 1 (EC 3.2.1.21).
GN BGL.
OS Talaromyces emersonii.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Talaromyces.
OX NCBI_TaxID=68825;
RN [1]
RP SEQUENCE FROM N.A.
RA Murray P.G., Collins C.M., Tuohy M.G.;
RT "Molecular cloning and expression analysis of beta glucosidase from
RL Talaromyces emersonii."
J. Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF439322; AAL34084.2; -.
DR EMBL; AY081764; AAL89551.2; -.
DR GO; GO:0004553; P:beta-glucosidase activity; IEA.
DR GO; GO:0004553; P:beta-glucosidase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 489 AA; 55810 MW; CF438E4ED72D52BA CRC64;

Query Match 49.1%; Score 1298; DB 3; Length 489;
Best Local Similarity 52.2%; Pred. No. 7.3e-96;
Matches 246; Conservative 70; Mismatches 149; Indels 6; Gaps 6;
```


Oy		247	WDSNEPRCKEAAERMEEHGIFNAPRIPFLKQDYPSPMKKOLGERLPALTTPADFAILNMG	306
Dd		255	WDPENPADVEACDKRIEPAISMFADPIHYGK-YIPSPMKQLDLDRPKPTPEETIAVH-GS	312
Oy		307	TDFYGMNYTTSQCFARHLHDG-EVPEDVDYGAIAHEOENKDGSFVGESGLAWLRSCPMFR	365
Dd		313	NDFPGMNHHCENYI.RNRTGEDPDE-DIAGNIIDLIMEDKNGNINIGETOCENLRPPPLGR	371
Oy		366	KHLARVYGLYCKP-IYTENGCPCCGEENNMTCEAVNDPFRIRYDSHLSISKAITODG	424
Dd		372	KLLKWLADRNVNPKIYVTENGTSYKGESDKPLEEVLANDEFVQVRYRDYIAGAVDAVADG	431
Oy		425	VVVGYFMALLDNLEMSDGYPRCGVFTDYTT-LKXTPKKSALVKDMF	474
Dd		432	VNVKAYMAWSLIDNFEWSEGRSRFGVIYVDIKNQCKRI PKXSALVIGELF	482
 RESULT 4 093785 PRELIMITARY; PRT; 466 AA.				
ID	093785			
AC	093785;			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
DE	Bete-glucosidase.			
CN	BGL2.			
OS	Trichoderma reesei (Hypocrea jecorina).			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.			
OX	NCBI_TaxID=51453;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=QM9414;			
RA	Takahima S., Nakamura A., Hidak M., Masaki H., Uozumi T.;			
RT	"Molecular cloning and expression of the novel fungal beta-glucosidase			
RL	genes from Humicola grisea and Trichoderma reesei.";			
J.	Biochem. 0:0-0(1999).			
DR	EMBL; AB003110; BAAV4959.1; --			
DR	HSSP; P22073; ITR1.			
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPRO01360; Glyco_hydro_1.			
DR	Pfam; PF00232; Glyco_hydro_1; 1.			
DR	.PRINTS; PR00131; GLHYDRLAS1.			
DR	PRODom; PD000650; Glyco_hydro_1; 1.			
DR	PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.			
SO	SEQUENCE 466 AA; 52240 MW; 73CAB66FCOA265B7 CRC64;			
 Query Match 48.0%; Score 1270; DB 3; Length 466; Best Local Similarity 52.7%; Pred. No. 1.2e-93; Matches 248; Conservative 65; Mismatches 142; Indels 16; Gaps 7				
Oy		7	LPNDEMFPATPAVOIEGAYVKGGRGPSIMWTYCHLEPRTNGANGDVACHYHRVDEDF	66
Dd		2	LPKDQMPATPAVAOIEGADVDDGGPSIMWTFCAQOPGKIAGSGSVTAICDSYNTRAEDI	61
Oy		67	DLLTKYGAAYAFSLSSWRSIIPLGRLDPVNBEGIEFYSKLIDALRRGITPWVLTYHW	126
Dd		62	ALLKSLGAKSRFSISWRSRIIPGGRGDVANQAGDIHYKVFDDLLDAGITTFITLFHM	121
Oy		127	LPQALHDIRYGMNLNVEVOLDBERYARLCFERFGDRVONMTINKPWICAIYGATGSA	186
Dd		122	LPEGIHQHYGGILLNTEPPLDFENYARVMF-FALPKVRMWITFNBPCLCSAIGYSGGT	180
Oy		187	PGRSSINGSTEGNATPEWMLGAKQAQNSHARAVALVYSMDPR-SQKQIGISLNGDYXE	245
Dd		181	PGROS-----TSEPWTYGHNILVAHGRAVAYRDPDFPASGGDIGIVLNGDFY	230
Oy		246	PWDSNEPRDKAEAREMEREFHIGFANPIFLKKQDYSPSMKKOLGERLPALTTPADFAILNAG	305
Dd		231	PWDADPADDKAEARELERLEFTAMFADPIYL-GDIYASPMKKOLGDRLPFTTPPERBALVH-G	288
Oy		306	ETDFYGMNYTTSQCFARHLHDGPVLETVDYGAIAHEOENKDGSFVGESGLAWLRSCPMFR	365

Db	SNDRYGNHNTYNSAIRHRSSPSADSDTVTGVNDVLFVNKGAGCIGETGSPMLRPPCAAGR	348
Db	289 SNDYGNHNTYNSAIRHRSSPSADSDTVTGVNDVLFVNKGAGCIGETGSPMLRPPCAAGR	348
Qy	366 KHLARVYGLVG-KRIYITENGCCPCGEBENMTCEAVNDPFRIRPDSHLDSIKAITDGG	424
Db	349 DFLWMSIKRKYGPPIYVTEGNTSJKGESDLPKEKILEDPPFKVYXNEYEIRAMVAVAEIDG	408
Qy	425 VVVGYPFAMALLDNLWSDGCGPRFGVFTDYTT-LKRTPKKSLVLKDMF	474
Db	409 VVVGYPFAMSLMDNFEWADGYVTRFGVTVYVDYENGQKRPFKKSALKPLF	459

RESULT 5

Q727X1 PRELIMINARY; PRT: 450 AA.

AC Q727X1;

DT 01-OCT-2003 (TREMblrel. 25, Created)

DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Beta-glucosidase 2.

OS BGL2.

OS Trichoderma viride.

OC Eukaryota; Fungi; Ascomycota; pezizomycotina; Sordariomycetes; Hypococreales; Hypocreales; mitosporic Hypocreales; Trichoderma.

OX NCBI_TaxID=5547;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AS 3.3711;

RA Liu B.D., Yang Q., Zhou Q.;

RT "Cloning and Sequence Analysis of the beta-glucosidase 2 Gene from Trichoderma viride AS 3.3711";

RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY343988; AAC21384.1; -.

SQ SEQUENCE 450 AA; 50020 MW; 9C945A16ECF288B2 CRC64;

Query Marce 44.3%; Score 1172; DB 3; Length 450;

Best Local Similarity 49.9%; Pred. No. 8.9e-85;

Matches 236; Conservative 65; Mismatches 132; Indels 40; Gaps 10

Qy	7 LPNDFEWGFATAVAQIEGAVKEGGRGSPIMDTYCHLBSRTNGANGVACDHYHRDEDF	66
Db	2 LPKDFQWGFATAVAQIEGAVDQDGRSPIMDTFCQPEKINDGSGVACDSYNTAEI	61
Qy	67 DLTTRYGAKAYRFSLSWRSRIIPLGRLDPVNEBEGIEFYSKILDALRRGITPMWTLHYMD	126
Db	62 ALLTSIGAKSYRFSISWSRIIPGGRGDAVNAQAGIDHYKKEVNDLDAGITPTFILFWM	121
Qy	127 LPQALHPYGGWLVVEVQVLDPEERYARLCFERFGDRVONMTINKPMTOATYGVATGSA	186
Db	122 LPQGLHGYGGWLVNTEPRLDFENYARWRFALPKVNMWTFNEPLCSALPGYSGGFA	180
Qy	187 PGRSSINRGSTEGTATEPMLAGKAQIISHARAVALVSRDPR-SQKQIGISLNGDYE	245
Db	181 PGRGS-----TSEPTVGNINILVAHGRAVAYNDPKPASGDQIGIVLNGDFTY	230
Qy	246 PWDNSNEPRDEKAERMEFHIGWPNAPIFLKKDYPESKKOLGERLPLTPADPAIILNAG	305
Db	221 PWDADPADKAEARERLEFFTAMPADPIYL-GDYPASMRKOLGRLPFTPEERALVH-G	288
Qy	306 ETDPEGMWYYSQFARHLDPVPEPDVIGALHENOENCDGSPVGEESGLAWLRSGPDNFR	365
Db	289 SNDYGNHNTYNSAIRHRSSPSADSDTVTGVNDVLFVNKGAGCIGETGSPMLRPPCAAGR	337
Qy	366 KHLARV-----YGLYKGPPIYITENGCCPCGEBENMTCEAVNDPFRIRYEDSHD	414
Db	338 --LAPVRRAATSMWCSAGTY-PPYIYITENGNTSJKGESDLPKEKILEDPPFKVYXNEYIR	394
Qy	415 SISKAITDGVVVGYPFAMALLDNLWSDGCGPRFGVFTDYTT-LKRTPKKS	466
Db	395 AMVTAVELDGVVVGYPFAMSLMDNFEWADGYVTRFGVTVYVDYENGQKRSPPRA	447

RESULT 6

Q9SP9 PRELIMINARY; PRT; 540 AA.
ID Q9SP9
AC Q9SP9
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Raucallicine-O-beta-D-glucosidase (EC 3.2.1.125).
OS Rauwolfia serpentina (Serpentwood) (Devilpepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Gentianales; Apocynaceae; Rauwolfiaceae; Vincetaceae; Rauwolfia.
OC NCBI_TaxID=4060;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BENTH. ex KURZ;
RA Warzcha H., Gerasimenko I., Kutchan T.M., Stoeckigt J.;
RT "Cloning, bacterial expression, and properties of raucallicine
RT glucosidase, an enzyme of indole alkaloid biosynthesis in Rauwolfia
RT serpentina";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF149311; AAF03675.1; -.
DR HSSP; P26205; 1CBG.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR PRODOM; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 540 AA; 60934 MW; 543A24654A4D5E7D CRC64;

Query Match 41.1%; Score 1087; DB 10; Length 540;
Best Local Similarity 45.3%; Pred. No. 7.9e-79;
Matches 226; Conservative 69; Mismatches 162; Indels 42; Gaps 14;
QY 8 PNDPEMGATTAAYOIEGAVKEGGRPSIMPTTYCHLEPSR-INGANGDVACHYHYRDEDF 66
DB 23 PADFMGCGSSAYOIEGARDGGRPSIMPTTYCHLEPSR-INGANGDVACHYHYRDEDF 82
QY 67 DLLTYGAKAYRFSISMSRIIPLGRLD-PVNEBGEFFYSKIDALLRGITPWTLYHW 125
DB 83 NILKMLGDAYRFSISMSRIIPLGRLD-PVNEBGEFFYSKIDALLRGITPWTLYHW 141
QY 126 DLPOALHRYGGLNVEEYQDLDFERYALCFERFGDRVQNMWITINXPMIOAIYATGSN 185
DB 142 DVPOALEDEYGGFUSPRIVD-DFCEYALCFMERFGDRVQNMWITINXPMIOAIYATGSN 200
QY 186 AGRSSIN----KST-----EGNTATEPMLAGKAOIMSHARAVAVSRD 226
DB 201 AGRGRTSPRHVNHPTVQHRCSYVAPQCICSTGNPTEFPYVTHHLLAHAAVELYKXK 260
QY 227 FRBQKQGTIGSLNGDIYEPWDSNEPRDKAERMEFHIGWANPIFLKQDYPSMKKQ 286
DB 261 FQROEGGIGISHATYQWMEPWDSNASVEAARALDFMLGMEVEPI-TSGDGYKSKKF 319
QY 287 LGEERLALTPDPAIINAGETDFGMYNTYSQFARHLDGVPVEDIYGA-----IH- 337
DB 320 VGSLLPFRSPQSKMLK-GSYDFVGLNYTASVYTN-----ASTNSSGSSNNFSYNDIHV 373
QY 338 EHQENKQSPVGEESGLAWLNSCPDMFRKILARVYGLYGRP-IYITENGCPCEGEENMTC 396
DB 374 TYETDRNGVPIPGSGSDMLIYPEGIRKLIVYTKKTYNNPLIYVTENGVDVAKNNLTL 433
QY 397 BEAANDPFRIRYPSHLDLSISKATYQCGVYVYGYFAMALLDNLEMSDGYPRFGVTFDDY 456
DB 434 SEARKDMLKLYLODHIENVRQAM-NDGVYVYGYFAMSLDNFEMGEGYVRRFGIITHIDY 492
QY 457 T-TLKRTPKKSALVTKDMF 474
DB 493 NDNFARYPRKDSAVWLMSF 511

RESULT 7
ID Q8L7J2 PRELIMINARY; PRT; 521 AA.
AC Q8L7J2
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Beta-glucosidase.
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euharoidae; Oryzaceae; Oryza.
OC NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Yukihikari;
RA Akiyama T., Ogasiri R., Ketudat-Cairns J.R., Swasti J., Been A.;
RT "Predominant expression of beta-glucosidase in germinating rice
RT seeds";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY129284; AA001354.1; -.
DR Grimeire; Q8L7J2; -.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR PRODOM; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
SQ SEQUENCE 521 AA; 58539 MW; 0554435869F3BD21 CRC64;

Query Match 39.7%; Score 1049; DB 10; Length 521;
Best Local Similarity 45.1%; Pred. No. 8.6e-76;
Matches 214; Conservative 79; Mismatches 157; Indels 24; Gaps 10;
QY 6 ALPNDPEMGATTAAYOIEGAVKEGGRPSIMPTTYCHLEPSRTNGANGDVACHYHYRDEDF 65
DB 49 SPEFGVFGTASAAAYOIEGAVKEGGRPSIMPTTYCHLEPSRTNGANGDVACHYHYRDEDF 108
QY 66 FDLTYGAKAYRFSISMSRIIPLGRLD-PVNEBGEFFYSKIDALLRGITPWTLYHW 125
DB 109 IQLMADMGDAYRFSISMSRIIPLGRLD-PVNEBGEFFYSKIDALLRGITPWTLYHW 166
QY 126 DLPOALHRYGGLNVEEYQDLDFERYALCFERFGDRVQNMWITINXPMIOAIYATGSN 185
DB 167 DLPOALEDEYKKMLBRLQVD-DFAAVALCFERFGDRVQNMWITINXPMIOAIYATGSN 225
QY 186 AGRSSIN----STEGNTATEPMLAGKAOIMSHARAVAVSRDPRSQKQIGISLNGDY 243
DB 226 AGRGCSVLHLHYCKAGNSCTEPEYVVAHHFILLHAAASIRTKYKATONGOLGIAFDVWM 285
QY 244 YEPWDSNEPRDKAERMEFHIGWANPIFLKQDYPSMKKQGERLPAITPADFALIN 303
DB 286 FRPM-SNTTIDLEAKRAQDFOLGMPADPFF-GDYPMRPRVGERLPRFT-ADEAIV 342
QY 304 AGETDFYGNNTYSQFARHLDGVPVEDIYGAIHQO-----EKKDSSPVGEESGL 354
DB 343 KQALDFVGVGHNTYTYTH----NNTWILGTLNNTLADTGIVSLPRKNGKPIGDRANS 397
QY 355 AMLRSCDPMFRKILARVYGLYGRP-IYITENGCPCEGEENMTCBEAVNDPRIRYPSHL 413
DB 398 IWLTVPRGMSLWNYVERNSPPVYITENGMDSNPFIIXDALDSKRIKYHNDYL 457
QY 414 DSIKATQDGVYVYGYFAMALLDNLEMSDGYPRFGVTFDDY-TLKRTPKKS 466
DB 458 TMLAASIKEDGDGVYGYFAMSLDNFEMGEGYVRRFGIITHIDY 511
RESULT 8
ID Q9P456 PRELIMINARY; PRT; 434 AA.
AC Q9P456
DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)
 DE Beta-glucosidase precursor.
 GN BGL1.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 NCBI_TaxID=5061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AMS11;
 RA Al Y.-C., Meng F.-M.;
 RT "Molecular cloning and sequencing of a beta-glucosidase gene from
 RT Aspergillus niger AMS11."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF269911; AAF74209.2; -.
 DR HSP; P26205; ICBG.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001360; Glyco_hydro_1.
 DR Pfam; PF00232; Glyco_hydro_1.
 DR PRINTS; PR00131; GHYDRLASE1.
 DR ProDom; PD000650; Glyco_hydro_1; 1.
 DR Signal.
 KW SIGNAL.
 FT CHAIN
 FT SIGNAL
 SQ SEQUENCE 434 AA; 48363 MW; 121ABE23AF2B78C CRC64;

Query Match 39.4%; Score 1042; DB 3; Length 434;
 Best Local Similarity 47.6%; Pred. No. 2,4e-75;
 Matches 224; Conservative 59; Mismatches 140; Indels 48; Gaps 13;

QY 7 LPNDFEMGFATYAOIEGAVGEGRPSIMDTYCHLEPSRTNGANGVACDHYRVEDF 66
 DB 2 LPNDLQGFKAQAQIEGAVDQGRPSIMDTFCAGQKTAQSSGVYACDSTNRVADI 61
 QY 67 DLTTKYGAAYRPSLSWSRIIPGRLDPVNEEGIEFYSKLIDALLRRGITPWVTLVHM 126
 DB 62 ALKSLGAKSYRPSIS-SR-IPBGRGDVANOAGIDHYVAFVDLLAGITPFTLPHMD 119
 QY 127 LPQALHRYGGMVNEVQDFFERYALCFERFGRVQNNITNXPMIOAIYGAATGSA 186
 DB 120 L--LHORVGLINRTFPLDFENYARVME-RALPKVRNM---NEPLCSAIPGVSGSFA 172
 QY 187 PGSSSINKSTEGTATEPMLAGKAQIMSHARAVVSRDPR-SQGGQIGISLNGEY 245
 DB 173 PGROS-----TSEPWTGNNILVAHGRAVAKAYRDFKPAASGGQIGIVLNGDFTY 222
 QY 246 PMDSNEPRDKEAERMEFIHGFANPIFLKDYPSMSKKOLGERLPALTPADFAILNAG 305
 DB 223 PMDAADPADKE-----RLFFTFAMPADPIYL-GDYPASMRKQDRLFTFPERALVH-G 276
 QY 306 ETDFFYGMVYTSQFARHLDPVPETDYGA-IHSHQENKDGSPVGEESGLAMLRSCPMFR 365
 DB 277 SNDFYGMNHTSYNIRHSSPASADTVGNVDVLFTHKQNCIGPFTQSPMLRCAAGFR 336
 QY 366 KHLARVYGLYG-KPIYITENGCCPGSENMTCBAVNDPRIRYFSDHLSISAKITQDG 424
 DB 337 DFLVMTSKRGSPIYITENGTSIKGSSDLPEKXITLDDPRVKKYNEYIAMVAVAVELDG 396
 QY 425 VVVGKGFAMALLDNLEWSDGYPFGVTFDTYTT-LKRTPKSAALVLCQNF 474
 DB 397 VNV-----RFGVTVVDYENGQGRFRPKSAKSLKPLF 427

RESULT 9
 Q9FIM4 PRELIMINARY; PRT; 490 AA.
 AC Q9FIM4;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)
 DE Beta-glucosidase.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=9915623; PubMed=10048486;
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen
 RT physically assigned P1 and TAC clones."
 RL DNA Res. 5:379-391(1998).
 DR EMBL; AB016877; BAB1630.1; -.
 DR HSP; P26205; ICBG.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001360; Glyco_hydro_1.
 DR Pfam; PF00232; Glyco_hydro_1; 1.
 DR PRINTS; PR00131; GHYDRLASE1.
 DR ProDom; PD000650; Glyco_hydro_1; 1.
 DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
 SQ SEQUENCE 490 AA; 56077 MW; 50DEB68FBF06D538 CRC64;

Query Match 39.2%; Score 1037.5; DB 10; Length 490;
 Best Local Similarity 45.0%; Pred. No. 6,6e-75;
 Matches 209; Conservative 83; Mismatches 155; Indels 17; Gaps 9;

QY 8 PNDPEMFATYAOIEGAVGEGRPSIMDTYCHLEPSRTNGANGVACDHYRVEDFD 67
 DB 22 PSFTFTGVATSAQIEGMMGKKGPSIMDKFTHIEGKLIDSGNDVADHYRKYKDVD 81
 QY 68 LTRKYGAAYRPSLSWSRIIPGRLDPVNEEGIEFYSKLIDALLRRGITPWVTLVHM 127
 DB 82 LIQQLGFGAYRFSISWSRIIP-DGLGTEVNEEGIAFYNDLINTLLEKGIOPYVTLVHMDL 140
 QY 128 PQALHRYGGMVNEVQDFFERYALCFERFGRVQNNITNXPMIOAIYGAATGSA 187
 DB 141 PSHLQEAIGGWTNRKLVY-FGLYADACFANFDRVHMTTLNEPLOTSVNGCIGIFAP 199
 QY 188 GRSSSINKSTEGTATEPMLAGKAQIMSHARAVVSRDPRSPQKQIGISLNGEY 247
 DB 200 GRNB-----KPLIEPIVSHHQLAHATANSIRSKYKESQGGQIGISVCEMAEP- 250
 QY 248 DSNEPRDKEAERMEFIHGFANPIFLKDYPSMSKKOLGERLPALTPADFAILNAGET 307
 DB 251 NSKPEDEKVAADRIRIDFGQMFDPPLFF-GDYPASMRKQDRLFTFPERALVH-G 309
 QY 308 DFFYGMVYTSQFARHLDPVPETDYGA-IHSHQENKDGSPVGEESGLAMLRSCPMFR 365
 DB 310 DFLGLNHTYSRLISHVSNKEAESNFYOQDELRIVELENDLLGERAASDMLAVPWGIR 369
 QY 366 KHLARVYGLYG-KPIYITENGCCPGSENMTCBAVNDPRIRYFSDHLSISAKITQDG 424
 DB 370 KTLNYSKTYNHPPIITENGMDDEDDGSASHMDLDDKRVVYFYSYLANVQAI-EDG 428
 QY 425 VVVGKGFAMALLDNLEWSDGYPFGVTFDTYTT-LKRTPKSA 467
 DB 429 VDIKGFAMSLDNLEWAGYTRFGVYVDYKGLTRHPSKA 472

RESULT 10
 Q8LF56 PRELIMINARY; PRT; 498 AA.
 AC Q8LF56;
 DT 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)
 DE Beta-glucosidase, putative.
 OC Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Db 273 FEP-ESKRTEDIEAQAQADPQGLGFLDPLMF-GDYPSMKSRVGRSLPVFTGSQSSLVK 330
QY 304 AGETDFGNNYTSQFARHLDPVETDYLGAIEHOENKDG-----SPGEESSL 354
Db 331 -GSLDFGNGHHTTYVARN-----NATNLGLTLHADVSSGTYLTFPKGLSTIGDAASS 384
QY 355 AMLRSCPMFRKHLARVGLYGR-PIYITENGCPCPEENMTCEAVNDPFRIRYFDSHL 413
Db 385 IWLIVYPRGRSLMNYIKHRYGNPPVITENGMDPNSILSRDALKDAKRIKHHDYL 444
QY 414 DJSKATODGVVVGKFAWALLDNLMSGCGYGRFVTTDY-TLKRTPKKSA 466
Db 445 SSLQASTIKEDGCVNGKGFVMSLLDNEMWAGYSSRFGLYFVDYRDLKRYPKDS 498
RESULT 12
QY073 PRELIMINARY; PRT; 489 AA.
AC QY073; (TEMBLrel. 25, Created)
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Latex cyanogenic beta glucosidase.
OS Hevea brasiliensis (Para rubber tree).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside 1; Malpighiales; Euphorbiaceae; Crotonoideae; Mitrandoeae;
OC NCBI_TaxID=3981;
RN [1]
RP SEQUENCE FROM N.A.
RA Sunderasan E., Yeang H.Y.;
RT "Hevea brasiliensis latex cyanogenic glucosidase (Hev b 4 a major latex allergen)".
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY297039; A051059.1; -
SQ SEQUENCE 489 AA; 56132 MW; 1D79CE651976F894 CRC64;
Query Match 38.9%; Score 1028.5; DB 10; Length 489;
Best Local Similarity 45.9%; Pred. No. 3.5e-74;
Matches 214; Conservative 83; Mismatches 146; Indels 23; Gaps 13;
QY 8 PNDPFWGFATAAYOIEGAVKEGGRGSPIMDYCHLEPSRTNGANGDVA CDHYHRVDEDFD 67
Db 24 PPFPLFGVATSAQIEGCGREGGRGSIWAFSHTKTIIIDGSGDVAVDHYRKYEDIE 83
QY 68 LTTYGAKAVRSLWSRIIP-LGGRLDVNEEGIEFYSKLIDALLRGITPWTLYHW 125
Db 84 LIAKLFDAVRFSLSWSRIIPDGLGTR--VNEGLAFYNNIINALLEKIEPYITLYHW 140
QY 126 DLPOALHDRVGMLANEBVQDLPERYARLCFERFGDGVONMTINXPMIOAIYGYATGSN 185
Db 141 DLPLHLQESNGWLNKEIVRY-FAIYADTCFASFGGRVKKMIIINLEPLOTAVANGPDTGL 199
QY 186 AGRSSINXKSTEGNTATEPMLAGKQIMSHARAVAYSRDPFSOKGQIGISLNGDYE 245
Db 200 APOK---HEIS-----YTFPLASHHQIILAHATVSIYRSMYDNQGGVGLVVDCEWAB 251
QY 246 PWDNSNEPRDKEAERMEFHIWFPANDIFLKQY PESHKKQLGERLPALTPADPAIINAG 305
Db 252 S-NSDKIEDKAAAKRLEFOLGMYLHPLY-GDYPERMKIIGGLGPKFSEEDKELLR-N 308
QY 306 ENDPYGMNYYTSQFAHLDGPVETDYLGA--IHEOENKDS PVGEESGLALRSCPM 363
Db 309 SLDFGNGHHTTYVARN-----NATNLGLTLHADVSSGTYLTFPKGLSTIGDAASS 384
QY 364 FRKHLARVGLYGR-PIYITENGCPCPEENMTCEAVNDPFRIRYFDSHLDSKAITO 422
Db 369 LRKLVNIYVGRVNNPIIYITENGMD-DESSAPLHEHLDKLAVRIFKGYLAANAQAI-K 426
QY 423 DGVVVGKFAWALLDNLMSDGYGRFVTTDY-TLKRTPKKSA 467

Db 427 DGADVRGYFAWSLLDNFEMAGYTKRFGILYVDYKNGLARHPKSSA 472
RESULT 13
QY0406 PRELIMINARY; PRT; 515 AA.
AC QY0406;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Putative beta-glucosidase.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriarthoideae; Oryzaceae; Oryza.
OC NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nipponbare;
RA Nacimento L., Spiegel L., de la Bastide M., Kirchoff K., King L.,
RA Preston R., Vil M.D., Baker J., Zucavern T., Santos L., Bell M.,
RA Miller B., Kuit K., Rodriguez S., Cunniff D.M., Balija V., Shah R.,
RA Bahret A., Palmer L., Yang C., O'Shaughnessy A., Dedhia N.,
RA McCombie W.R.;
RT "Genomic Sequence for Oryza sativa, Nipponbare strain, clone
RT OJNBA0055C16, from chromosome 10, complete sequence."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC074354; AK92581.1; -
DR Gramine; Q94H06; -
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR Interpro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLYCDRLASE1.
DR Prodom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
SQ SEQUENCE 515 AA; 58683 MW; B842BD29627A8BC CRC64;
Query Match 38.7%; Score 1023; DB 10; Length 515;
Best Local Similarity 42.3%; Pred. No. 1e-73;
Matches 208; Conservative 88; Mismatches 160; Indels 36; Gaps 12;
QY 6 ALPNDPFWGFATAAYOIEGAVKEGGRGSPIMDYCHLEPSRTNGANGDVA CDHYHRVDEDFD 65
Db 37 SPFNGVFGTASSAYIEGAVKEGGRGSPIMDYCHLEPSRTNGANGDVA CDHYHRVDEDFD 96
QY 66 FDLITYGAKAVRSLWSRIIP-LGGRLDVNEEGIEFYSKLIDALLRGITPWTLYHW 125
Db 97 IQLMADMGMDAVRFSLSWSRIIPNG--TGEVNOAGIDHYNKLIINALLAGIEPYITLYHW 154
QY 126 DLPOALHDRVGMLANEBVQDLPERYARLCFERFGDGVONMTINXPMIOAIYGYATGSN 185
Db 155 DLPOALHDRVGMLANEBVQDLPERYARLCFERFGDGVONMTINXPMIOAIYGYATGSN 185
QY 155 DLPOALHDRVGMLANEBVQDLPERYARLCFERFGDGVONMTINXPMIOAIYGYATGSN 185
Db 155 DLPOALHDRVGMLANEBVQDLPERYARLCFERFGDGVONMTINXPMIOAIYGYATGSN 185
QY 186 AGRSSINXKSTEGNTATEPMLAGKQIMSHARAVAYSRDPFSOKGQIGISLNGDYE 243
Db 214 AGRSGVLLHLYCKKNSGTEPIYVAHMTLSHATYSIDYRKKYKASQNGELISDVW 273
QY 244 YEPWDSNEPRDKEAERMEFHIWFPANDIFLKQY PESHKKQLGERLPALTPADPAIINAG 303
Db 274 YEPWDSNEPRDKEAERMEFHIWFPANDIFLKQY PESHKKQLGERLPALTPADPAIINAG 303
QY 304 AGETDFGNNYTSQFAHLDGPVETDYLGA--IHEOENKDS PVGEESGLALRSCPM 363
Db 332 -GSLDFGNGHHTTYVARN-----NATNLGLTLHADVSSGTYLTFPKGLSTIGDAASS 384
QY 355 AMLRSCPMFRKHLARVGLYGR-PIYITENGCPCPEENMTCEAVNDPFRIRYFDSHL 413
Db 386 IWLIVYPRGRSLMNYIKHRYGNPPVITENGMDPNSILSRDALKDAKRIKHHDYL 444
QY 405 RIRYFDSHLDSKAITODGVVVGKFAWALLDNLMSDGYGRFVTTDY-TLKRTPKKSA 466
Db 442 RTKYHNDYLTNLADSIREDGCVRGYFAWSLLDNEMWAGYSSRFGLYFVDYRDLKRYPKDS 498

Qy 465 KSALVLKDMFAA 476
Db 502 NSVQWFKNLLAS 513

RESULT 14
ID 07XFUB PRELIMINARY; PRT; 515 AA.
AC 07XFUB; 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Putative beta-glucosidase.
GN OSJNBA065C16.15.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RT chromosome 10."
RL Science 300:1566-1569(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Bueli C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB017075; AAP52953.1; -
SQ SEQUENCE 515 AA; 58683 MW; B842BD29627AE8CC CRC64;

Query Match 38.7%; Score 1023; DB 10; Length 515;
Best Local Similarity 42.3%; Pred. No. 1e-73;
Matches 208; Conservative 88; Mismatches 160; Indels 36; Gaps 12;

Qy 6 ALPNDFFGATTAAYOIEGAVKEGGRGSPIMDYCHLEPSRTNANGVACDHYRVED 65
Db 37 SFGPGFVGITASSAYOEGAVKEGGRGSPITWDKPAHFGKIIDPSNAVAADQYHREED 96
Qy 66 FDLTLTKGAKAYRPSLSWSRIIPLGRLDPVNEBGEIEFYSLIDALLRGITPWTLYHM 125
Db 97 IQLMADKMDAYRSISWSRIIFPNC--TGEVNOAGIDHYNKLINALAKGIEPYLYHM 154
Qy 126 DLPOALHRYGWLINVEEVLDFERYARLCEFRGDRVQNMITTINXPMIOAIYGATGSN 185
Db 155 DLPOALEDKYTGWLD-RQINDYAVYAEYTCFOAFGDRVKHWITENEPHTVAOVAVDSGMH 213
Qy 186 APGSSINIKH--STEGNATEPWLAKGAQIMSHARAVALVSRDPRPSQKQIGISLNDY 243
Db 214 APGRGSVLLHYCKKNGSTEPYIVAHNMILSHAVSDIYRKRYASONGELISFDVIT 273
Qy 244 YEPMDNSPPRKEAERMEFHIGFANPIFLKDYPSMKKOLGERLPALTPADFAILN 303
Db 274 YEPM-SNSTADIEAKRAQEFQLGWADPFF-GYPRATMSRVSRLPKTEKKAALVN 331
Qy 304 AGETDFYGMNYTTSQFAH-----LDGPVPEPTDYLGAIIHEHOKNKGSPVGEESGL 354
Db 332 -GSLDFMGINHYTFTYKDDQSTVIEKLINTLTADTATISV-----PPRNGQPIGDRANS 385
Qy 355 AMVSSCPMFRKHLARVYGLYKPR-IYITENGCCPGRBENMTCE-----EAVNDPF 404
Db 386 IWLIVTPSRKMLNMYVXDRNKPFTVYITENG-KC--TYVICDLFPFISIKNLLKODK 441
Qy 405 RIRFEDSHLDSISKAITODGVVVKSGYFAMALLDNLEMSDGYGPRGVTFTDYLTKRTPX 464
Db 442 RTKYHNDILTNLADSIREDGCDVRGYPFAMSLDNNEMAGATSRGLVYVYKKNKRYPX 501
Qy 465 KSALVLKDMFAA 476

Db 502 NSVQWFKNLLAS 513

RESULT 15
ID 094513 PRELIMINARY; PRT; 511 AA.
AC 094513; 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE Prunasin hydrolase isoform PH A (EC 3.2.1.116) (Fragment).
GN PH-L2.
OS Prunus serotina (Black cherry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eumossids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX NCBI_TaxID=23207;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhou J., Hartmann S., Shepherd B.K., Poulton J.E.;
RT "Investigation of the microheterogeneity and aglycone specificity-
RT conferring residues of black cherry (Prunus serotina) prunasin
RT hydrolases."
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF413214; AAL07435.1; -
DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR Interpro: IPR001360; Glyco_hydro_1.
DR Pfam: PF00232; Glyco_hydro_1; 1.
DR PRINTS: PR00131; GLYDRLASE1.
DR ProDom: PD000650; Glyco_hydro_1; 1.
DR PROSITE: PS00653; GLYCOSYL_HYDROL_Fr_2; 1.
KM Glycosidase; Hydrolase.
FT NON TER 1
SQ SEQUENCE 511 AA; 58236 MW; 9798674EB4C9174 CRC64;

Query Match 38.5%; Score 1017; DB 10; Length 511;
Best Local Similarity 43.8%; Pred. No. 3.1e-73;
Matches 213; Conservative 82; Mismatches 167; Indels 24; Gaps 12;

Qy 11 FEMGFATTAAYOIEGAVKEGGRGSPIMDYCHLEPSR-TNANGVACDHYRVEDDL 69
Db 21 FIFGTASAAAYOIEGAVKEGGRGSPIMDYTHNHPERIKDMSGDIALDQYHREVDGIM 80
Qy 70 TKYGAAYRPSLSWSRIIPLGRLDPVNEBGEIEFYSLIDALLRGITPWTLYHM 129
Db 81 KKNGLDYLRLSISWSRLPRNGKLSGVNKGIEYNNLTMBLLRNGITPFTLPHMDVPQ 140
Qy 130 ALHRYGWLINVEEVLDFERYARLCEFRGDRVQNMITTINXPMIOAIYGATGSNAPGR 189
Db 141 ALVDEYGGFLSPRIVD-HYXDYTELCFKEFGDRIKHWITINPEYAVSHHGVAIGIHAAGR 199
Qy 190 SSINKHSTEGNTANPEPWLAKGAQIMSHARAVALVSRDPRPSQKQIGISLNDYEPWDS 249
Db 200 CSWDEACLGDSALIEPLVTHNQLAHSAVKYKDKYQASQNCVIGITVYVSHWIEP-AS 258
Qy 250 NEPRDKEAERMEFHIGFANPIFLKDYPSMKKOLGERLPALTPADFAILNAGETDF 309
Db 259 KSKEDIDAAKRYLDPMFGWMPSPPLTI-GDYPHSRHLVGERLPVFTBEQSKLTN-GSFD 316
Qy 310 YGMNYTTSQFAHLDGPVPEPTDYGA--IHEHOKN--KDSGPVGEESGLAMLRBCPD 362
Db 317 IGLNYGARVASDE-----SNDYIAPSPSYLIDHRANVTTELNGVPIPRGASDMLVYYPE 371
Qy 363 MFRKHLARVYGLYKPR-IYITENGCCPGRBENMTCEAVNDPFRIRFEDSHLDSISKAIT 421
Db 372 GIYLLLTHTKETYNPLIYITENGIIDEPNNPKSLSEELNDYTRIIDYTHHLCYQAAI- 430
Qy 422 QDGVVVKSGYFAMALLDNLEMSDGYGPRGVTFTDYLTKRTPKSAVLKDMF-----AA 476
Db 431 KDGVRVKGYPFAMSLDNNEMAGATSRGLVYVYKKNKRYPX 501
Qy 477 RORVKV 482

Db 491 :::|
 KKEIRV 496

Search completed: March 25, 2004, 15:56:36
Job time : 48 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2004, 15:45:56 ; Search time 17 Seconds
(without alignments)
1482.468 Million cell updates/sec

Title: US-10-026-140-2
Perfect score: 2644
Sequence: 1 MPESLALPNDPFWGFAATAY.....KSALVTKDMPARQVKA 484

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	880.5	33.3	493	1 BGLS_TRIRP	P26204 trifolium r
2	879	33.2	566	1 BGLC_MAIZE	P49235 brassica m
3	858.5	32.5	548	1 MYRO_BRANA	Q00336 brassica na
4	857.5	32.4	1927	1 LPH_HUMAN	P09848 homo sapien
5	852.5	32.2	1936	1 LPH_RABIT	P09849 corycolagus
6	851.5	32.2	1928	1 LPH_RAT	Q02401 rattus norv
7	849.5	32.1	541	1 MYRO_ARATH	P37702 arabidopsis
8	845	32.0	544	1 MYR3_SINAI	P29092 sinapis alb
9	839	31.7	446	1 BGLA_THEMA	Q08638 thermocoga
10	831	31.4	528	1 BGL_ARATH	Q08650 arabidopsis
11	809	30.6	444	1 BGLA_THENE	Q03843 thermocoga
12	802	30.3	449	1 BGLA_BACCI	Q03506 bacillus cl
13	791	29.9	448	1 BGLB_PABPO	P22505 paenibacilli
14	790	29.9	425	1 BGLT_TRIRP	P26205 trifolium r
15	789	29.8	448	1 BGLA_PABPO	P22073 paenibacilli
16	767	29.0	459	1 BGLS_AGRSP	P12614 agrobacteri
17	749	28.3	448	1 BGLA_CLODM	P26208 clostridium
18	729.5	27.6	455	1 BGLS_CALSA	P10482 caldocellum
19	702	26.6	477	1 BGL2_BACSU	P22403 bacillus su
20	689.5	26.1	473	1 BGLB_MICBI	P22403 bacillus su
21	614	23.2	468	1 BGLG_LACIA	P11546 lactococcus
22	607.5	23.0	474	1 LACG_LACCA	P11565 lactobacilli
23	599	22.7	470	1 LACG_STANU	P11175 staphylococ
24	597	22.6	470	1 LACG_STAM	Q09878 staphylococ
25	590.5	22.3	473	1 LACG_LACAC	P50978 lactobacilli
26	587	22.2	468	1 LACG_STRMU	P50978 streptococc
27	583	22.0	470	1 LACG_STAPD	Q09878 staphylococ
28	558.5	21.1	480	1 BGLA_ERWHE	Q09437 erwilia her
29	530.5	20.1	479	1 BGLA_BACSU	P22973 bacillus su
30	522.5	19.8	479	1 BGLA_ECOLI	Q46829 escherichia
31	497	18.8	473	1 ABGA_CLODO	Q46130 clostridium
32	491	18.6	464	1 CASB_XLBBOX	Q48409 klebsiella
33	485	18.3	470	1 BGLB_ECOLI	P11988 escherichia

34	475.5	18.0	474	1 ASCB_ECOLI	P24240 escherichia
35	474.5	17.9	469	1 BGLI_BACSU	P40740 bacillus su
36	470	17.8	465	1 ARBB_ERWCH	P26206 erwilia chr
37	358.5	13.6	489	1 BGLA_SULSO	P22498 sulfolobus
38	344.5	13.0	489	1 BGLA_SULSH	P50388 sulfolobus
39	343.5	13.0	491	1 BGLA_SULAC	P14288 sulfolobus
40	304.5	11.5	243	1 MYR2_SINAI	P29736 sinapis alb
41	302	11.4	248	1 MYR1_SINAI	P29737 sinapis alb
42	301.5	11.4	244	1 MYR1_SINAI	P29737 sinapis alb
43	97.5	3.7	1088	1 YMS4_YEAST	Q04336 saccharomyc
44	96.5	3.6	766	1 YMS4_YEAST	P57703 pseudomonas
45	96	3.6	876	1 SYL_RHIME	Q92Kw8 rhizobium m

ALIGNMENTS

RESULT 1	ID	BGLS_TRIRP	STANDARD	PRT	493 AA.
AC	P26204				
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DE	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	Non-cyanogenic beta-glucosidase precursor (EC 3.2.1.21)				
OS	Trifolium repens (Creeping white clover)				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids;				
OC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.				
NC	NCBI_TaxID=3899;				
RN	[1]				
RP	SEQUENCE FROM N.A. AND SEQUENCE OF 135-157.				
RC	STRAIN=5100 (EC); TISSUE=leaf;				
RX	MEDLINE=91322517; PubMed=1907511;				
RA	Oxley E., Dunn M.A., Pancoro A., Hughes M.A.;				
RT	"Nucleotide and derived amino acid sequence of the cyanogenic beta-				
RT	glucosidase (linamarase) from white clover (Trifolium repens L.).";				
RL	Plant Mol. Biol. 17:209-219(1991).				
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-				
CC	glucose residues with release of beta-D-glucose.				
CC	-I- TISSUE SPECIFICITY: Leaves.				
CC	-I- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X56734; CA40058.1; ..				
DR	PIR; S16581; GLY31.				
DR	HSSP; P26205; 1CBG.				
DR	InterPro; IPR001360; Glyco_hydro_1.				
DR	Pfam; PF00232; Glyco_hydro_1; 1.				
DR	PRINTS; PR00131; GLYTRASE1.				
DR	PRODom; PD000650; Glyco_hydro_1; 1.				
DR	PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1.				
DR	PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.				
KM	Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	CHAIN	1	18		POTENTIAL.
FT	CHAIN	19	493		NON-CYANOGENIC BETA-GLUCOSIDASE.
FT	ACT_SITE	204	204		PROTON DONOR (POTENTIAL).
FT	ACT_SITE	422	422		NUCLEOPHILE (BY SIMILARITY).
FT	CARBOHYD	34	34		N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	335	335		N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	371	371		N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	412	412		N-LINKED (GLCNAC...) (POTENTIAL).
SQ	SEQUENCE	493 AA;	55960 MW;	B6B3BA5B8FCC8F	CRCC4;
Query Match	33.3%;	Score 880.5;	DB 1;	Length 493;	
Best Local Similarity	40.3%;	Pred. No. 8.2e-61;			

```
Matches 186; Conservative 80; Mismatches 177; Indels 19; Gaps 8;
QY 6 ALPNDMEGFATPAAYOIEGAVKGGKPSIMPTVYCHLEPSR-TNGANGDVACHHARDE 64
DB 39 SFFRGRIIFAGSSAYOIEGAVNKGKPSIMPTVTHKPEKIRDSNADITVDQYHRYKE 98
QY 65 DFDLTKYAKAYRPSLSWSRIIPUGRLDPVNEEGIEFYSKLDALRRGITTPVTLVH 124
DB 99 DVGIMKQMDSDYRFISWPRILPKKLSGGINHGKIKYNNLWELLANGIQPVTILFH 158
QY 125 WDLPOLAHRYGGMVNEVEVQD-DFERYARLCFEREGDRQWMTTINXPMQAIYIATGS 184
DB 159 WDLPOLLEDEYGGFPLNSGVIN-DFPDYDULCFKEDRVRVMTSTNEPWFNSGVALGT 217
QY 185 NAGRSSINKHSTEGNTATEPWLAKAQMISHARAVAYSRDPRFSQKQIGISLNGDY 244
DB 218 NAGRCASANVAKPGDSGSGPIVTHNOILAHBAVHYKTYQAYQKGIITLVSNWL 277
QY 245 EFWDSNEPDKAARERNEFHIGWFPANPIFLKKDY PESNKKQGLRRLPADPAIINA 304
DB 278 WPLDONSIPDIKAERSLDFQGLFMEQL-TTGDSYKSMRRIYKRLPKFSFESSLVN- 335
QY 305 GETDPYGMVYTSOFAHLDG-----PVETDYLGAHHEHQENKQSGVYGESGLAMLR 359
DB 336 GSFDFICINYSSTISNAPSHGNAKPSYSTPMNTI--SFEKRGIFLGPRAASIMTYV 392
QY 360 CPDMFRKHLARVYLYGK-----PIYITENGCPGCEENMTCEAVNDPRIRYFDSHL 413
DB 393 YPWYFIQEDFEIFCYILKINTITLQFSITENGMEFNATLPVEBALNTYRIDYVYHL 452
QY 414 DSISAIATQDGVVVGYPAMALLDNLWMSDGYGPRFGVTFD 455
DB 453 YYIRSAI-RAGSNVKGAFYAMSLDCEMFAGFTVRFGLNFVD 493

RESULT 2
BGLC_MAIZE
ID_BGLC_MAIZE STANDARD; PRT; 566 AA.
AC P49235;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Beta-glucosidase, chloroplast precursor (EC 3.2.1.21) (Gentiobiase)
DE (Celllobiase) (Beta-D-glucoside gluconhydrolase).
GN GLUT1
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. INHERD LINE K55; TISSUE=Shoot;
RA Esen A.; Shahid M.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MUTIN; TISSUE=coleoptile;
RX MEDLINE=94053747; PubMed=8235622;
RA Brixobahy B., Moore I., Kristoffersen P., Bako L., Campos N.,
RA Schell J., Palme K.;
RT "Release of active cytokinin by a beta-glucosidase localized to the
RT maize root meristem";
RL Science 262:1051-1054(1993).
RN [3]
RP SEQUENCE OF 55-74, AND CHARACTERIZATION.
RC STRAIN=CV. INHERD LINE K55;
RA Esen A.;
RT "Purification and partial characterization of Maize (Zea Mays L.)
RT beta-glucosidase";
RL Plant Physiol. 98:174-182(1992).
RN [4]
RP SEQUENCE OF 55-69; 165-174; 207-213 AND 217-235.
```

```
RC TISSUE=Coleoptile;
RA Touzet P., Riccardi F., Morin C., Demerval C., Huet J.-C.,
RA Fernollet J.-C., Zavy M., de Vienne D.;
RT "The maize two dimensional gel protein database: towards an integrated
RT genome analysis program";
RL Theor. Appl. Genet. 93:997-1005(1996).
CC -1- FUNCTION: Is implicated in many functions such as ABA metabolism,
CC hydrolysis of conjugated gibberellins, conversion of storage forms
CC of cytokinins to active forms. Also acts in defense of young plant
CC parts against pests via the production of hydroxamic acids from
CC hydroxamic acid glucosides. Enzymatic activity is highly
CC correlated with plant growth. Its optimal pH is 5.8, and optimal
CC temperature 50 degrees Celsius. Activity is totally lost when the
CC enzyme is heated at 55 degrees Celsius.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC glucose residues with release of beta-D-glucose.
CC -1- ENZYME REGULATION: Reversibly inhibited by micromolar
CC concentrations of Hg(2+) or Ag(+), but irreversibly inhibited by
CC alkylation in presence of urea.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- TISSUE SPECIFICITY: Most abundant in the coleoptile.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; U25157; AA65946.1; -.
DR EMBL; X74217; CA55293.1; -.
DR PIR; A48860; A48860.
DR PDB; 1E1E; 19-FEB-01.
DR PDB; 1E1F; 19-FEB-01.
DR PDB; 1E4L; 11-DEC-00.
DR PDB; 1E4N; 11-DEC-00.
DR PDB; 1E5J; 11-DEC-00.
DR PDB; 1E5G; 11-DEC-00.
DR Maize-2DPAGE; P49235; COLEOPTILE.
DR MaizeDB; 13870; -.
DR Interpro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GHYDRASE1.
DR Prodom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Hydrolyase; Glycosidase; Chloroplast; Transit peptide; 3D-structure.
FT TRANSIT 1
FT CHAIN 55
FT ACT SITE 245
FT ACT SITE 460
FT CONFLICT 477
FT CONFLICT 551
FT CONFLICT 554
SQ SEQUENCE 566 AA; 64237 MW; 4EA241258AE3641B CRC64;

Query Match 33.2%; Score 879; DB 1; Length 566;
Best Local Similarity 38.3%; Pred. No. 1.3e-60;
Matches 192; Conservative 94; Mismatches 183; Indels 32; Gaps 14;
QY 1 MPESIALPNDMEGFATPAAYOIEGAVKGGKPSIMPTVYCHLEPSR-TNGANGDVACHH 59
DB 72 IPQRDWFPSDFTFGATATSAVOIEGAMNDGKESWMDHFCNHPRIIRIDSGNSY 131
QY 60 HRYDEDFDLTKYAKAYRPSLSWSRIIPUGRLDPVNEEGIEFYSKLDALRRGITPW 119
DB 132 HMYKTDVRLKEMGMDAVRFISWPRILPKKKEGINPDGKIKYRNILNLENGIRPY 191
QY 120 VTLVHMDPOLAHRYGGMVNEVEVQD-DFERYARLCFEREGDRQWMTTINXPMQAI 177
DB 192 VTFHMDVPOALEEKYGGFLDKSHKSLVEDYTYFAKCFDNFQDKVKNMLTENEPQETS 251
```

```

OY 178 YGVATGSAAPRRSS--INXSTSENTATEPEMLACKQAQMSHARAVAVYSDRFRSQGOI 235
Db 252 FSYGIGVAFAPRCSPGLDCAAPPTGNSLVEPTTAGHNILAHAEVADLVNKHAKDDT-RI 310
OY 236 GISLNGDYEEWQMSNEPRDXEAERMEFHIGWEFANPIFLKKDYPESNKKOLGERLPALT 295
Db 311 GLARDVWGVRPYGRSF-LDKQAERMSDINLQWTFLEPV-VARGDYPFMSRLARERLPPFK 368
OY 296 PADPAIINAGETDFYGMNYYTTSQPARHLD-----GPPETDYLGAIHEHQNKGSPPGE 350
Db 369 DEQEKEL-AGSYNNMGLNYYTSRFSKNIDISPNYSPLVNTDAAVASOE-VNGPDGKPIGP 426
OY 351 ESGLAWLRSQCDMRKHLARYGLYGYK-PIYITENG-----CPCGSENMTCCEAVN 401
Db 427 PMGPMPIYMPPEGLKDLMTMKAKRYGNPPIITENGIDVDUTKETPLD-----MEALN 480
OY 402 DPFRIREFDHLSDISKAITODGVVVKGYFAMALLDLNLEWSDGYGPREGVFTFD-YTTLK 460
Db 481 DYKRLDYIQRHIALTKESIDL-GSNVQGYFAMSLDNEFWAGFTERGIYVDRNNCT 539
OY 461 RTPKSAVLVKOMFARORVK 481
Db 540 RYMKESAKWLKEFNTAKKPSK 560

RESULT 3
MYRO_BRANA
AC 000326;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myrosinase precursor (EC 3.2.1.147) (Sintgrinase) (Thioglucosidase).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Svalofs Karat;
RA Falk A., Xue J., Jenman M., Raak L.;
RT "Sequence of a cDNA clone encoding the enzyme myrosinase, and
expression of myrosinase in different tissues of Brassica napus.";
RL Plant Sci. 83:181-186(1992).
CC -1- FUNCTION: Degradation of glucosinolates (glucose residue linked by
a thioglucoside bound to an amino acid derivative) to glucose,
sulfate and any of the products: thiocyanates, isothiocyanates,
nitriles, epithionitriles or oxazolidine-2-thione.
CC -1- CATALYTIC ACTIVITY: H(2)O + a thioglucoside = a sugar + a thiol.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Vacuolar.
CC -1- TISSUE SPECIFICITY: In vacuoles called myrosin grains of a certain
class of cells, myrosin cells, distributed in the cotyledons and
the axis of the embryo as well as in different organs of the
growing plant.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use by
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, X60214; CAA42775.1; -.
DR PIR, S26149; S26149.
DR HSSP, P26205; 1CRG.
DR InterPro, IPR001360; Glyco_hydro_1.
DR Pfam, PF00232; Glyco_hydro_1; 1.
DR PRINTS, PR00131; GLHYDRASE1.

```

Query Match	32.5%	Score 858.5	DB: 1	Length 548
Beet Local Similarity	40.1%	Pred. No. 4.8e-59		
Matches 194	Conservative	87	Mismatches 156	Indels 47
				Gaps 16
Dr. PRODOM: PD000650; GlycoHydrol. F1.1; 1.				
DR PROSITE: PS00572; GLYCOSYL_HYDROL_F1.1; 1.				
DR PROSITE: PS00653; GLYCOSYL_HYDROL_F2.1; 1.				
KW Hydrolyase; Glycosidase; Multigene family; Glycoprotein; Signal.				
FT SIGNAL	1	20	BY SIMILARITY.	
FT CHAIN	21	548	MYOSINASE.	
FT ACT SITE	429	429	NUCLEOSIDE (BY SIMILARITY).	
FT CARBOHYD	110	110	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CARBOHYD	240	240	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CARBOHYD	331	331	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CARBOHYD	520	520	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SEQUENCE	548 AA;	62735 MW;	F30346A01CD732A CRC64;	
Query Match	32.5%	Score 858.5	DB: 1	Length 548
Beet Local Similarity	40.1%	Pred. No. 4.8e-59		
Matches 194	Conservative	87	Mismatches 156	Indels 47
				Gaps 16
Dr. PRODOM: PD000650; GlycoHydrol. F1.1; 1.				
DR PROSITE: PS00572; GLYCOSYL_HYDROL_F1.1; 1.				
DR PROSITE: PS00653; GLYCOSYL_HYDROL_F2.1; 1.				
KW Hydrolyase; Glycosidase; Multigene family; Glycoprotein; Signal.				
FT SIGNAL	1	20	BY SIMILARITY.	
FT CHAIN	21	548	MYOSINASE.	
FT ACT SITE	429	429	NUCLEOSIDE (BY SIMILARITY).	
FT CARBOHYD	110	110	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CARBOHYD	240	240	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CARBOHYD	331	331	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT CARBOHYD	520	520	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SEQUENCE	548 AA;	62735 MW;	F30346A01CD732A CRC64;	

RT "Complete primary structure of human and rabbit lactase-phlorizin
RT hydrolase: implications for biosynthesis, membrane anchoring and
RT evolution of the enzyme.";
RM EMBL J. 7:2705-2713(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91206402; PubMed=1902057;
RA Bohl W., Wagner P., Mantel N.;
RT "Structure of the chromosomal gene and cDNAs coding for lactase-
RT phlorizin hydrolase in humans with adult-type hypolactasia or
RT absence of lactase.";
RL Am. J. Hum. Genet. 48:889-902(1991).
CC -1- FUNCTION: LPH splits lactose in the small intestine.
CC -1- CATALYTIC ACTIVITY: Lactose + H(2)O = D-glucose + D-galactose.
CC -1- CATALYTIC ACTIVITY: Glycosyl-N-acetylsphingosine + H(2)O = a sugar +
CC N-acetylsphingosine.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Brush border.
CC -1- TISSUE SPECIFICITY: Intestine.
CC -1- DOMAIN: THE SEQUENCE EXHIBITS 4 REGIONS (I-IV) OF INTERNAL
CC HOMOLOGY; THEREFORE LPH MIGHT HAVE EVOLVED BY TWO CYCLES OF
CC PARTIAL GENE DUPLICATION.
CC -1- DISEASE: Defects in the upstream vicinity of LCT are the cause of
CC disaccharide intolerance III [MIM:223100]; also known as adult
CC lactase deficiency or adult-type hypolactasia. In many human
CC populations the activity of LCT declines in adults, leading to
CC adult-type hypolactasia, whereas in other populations the high
CC activity persists.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
CC or send an email to license@ebi.ac.uk).

DR EMBL; X07994; CAA30801.1; -;
DR EMBL; M61850; AAA59504.1; -;
DR EMBL; M61834; AAA59504.1; JOINED.
DR EMBL; M61835; AAA59504.1; JOINED.
DR EMBL; M61837; AAA59504.1; JOINED.
DR EMBL; M61838; AAA59504.1; JOINED.
DR EMBL; M61839; AAA59504.1; JOINED.
DR EMBL; M61840; AAA59504.1; JOINED.
DR EMBL; M61841; AAA59504.1; JOINED.
DR EMBL; M61842; AAA59504.1; JOINED.
DR EMBL; M61843; AAA59504.1; JOINED.
DR EMBL; M61844; AAA59504.1; JOINED.
DR EMBL; M61845; AAA59504.1; JOINED.
DR EMBL; M61846; AAA59504.1; JOINED.
DR EMBL; M61847; AAA59504.1; JOINED.
DR EMBL; M61848; AAA59504.1; JOINED.
DR EMBL; M61849; AAA59504.1; JOINED.
DR PIR; S01168; S01168.
DR HSSP; P26205; 1CBG.
DR GeneW; HGNC:6530; LCT.
DR MIM; 603202; -;
DR MIM; 223100; -;
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0000016; F:lactase activity; TAS.
DR InterPro; IPR01360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 4.
DR PRINTS; PR00131; GLHYDLASL.
DR ProDom; PD000650; Glyco_hydro_1; 4.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 2.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 3.
KW Hydrolase; Glycosidase; Zymogen; Signal; Transmembrane; Repeat.
FT SIGNAL 1 19
FT PROPEP 20 866 BETA-GLUCOSIDASE.
FT CHAIN 867 1927 LACTASE-PHLOIRIZIN HYDROLASE.

FT DOMAIN 20 1882 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 1883 1901 POTENTIAL.
FT DOMAIN 1902 1927 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 87 1841 4 X APPROXIMATE REPEATS.
FT REPEAT 87 172 1.
FT REPEAT 87 172 2.
FT REPEAT 87 172 3.
FT REPEAT 87 172 4.
FT ACT_SITE 1370 1841 4.
FT ACT_SITE 1065 1065 PROTON DONOR (POTENTIAL).
FT ACT_SITE 1273 1273 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 1538 1538 PROTON DONOR (POTENTIAL).
FT ACT_SITE 1749 1749 NUCLEOPHILE (BY SIMILARITY).
FT CONFLICT 219 219 V -> I (IN REF. 2).
FT CONFLICT 1639 1639 N -> S (IN REF. 2).
SQ SEQUENCE 1927 AA; 218601 MW; FA748640B9A35A6F CRC64;

Query Match 32.4%; Score 857.5; DB 1; Length 1927;
Best Local Similarity 40.3%; Pred. No. 3e-58;
Matches 192; Conservative 75; Mismatches 167; Indels 43; Gaps 12;

QY 8 PNDEMGFATAVAOIEGAVKEGRGSPISIMPTCYHLEPSRTNGANGVACDHYRDEPD 67
Db 1378 PEGFTWSASAAVAQIEGAWRADGKLSIMPTFSHTPLRVENDAGVACDHYRIADLV 1437
QY 68 LITTYGAKAVRFLSWSRIIPLGRLDPVNEEGIEFYSKLIDALRRGITPWTLYHMDL 127
Db 1438 TLQNLGVSHYRFSISMSRIIP-DGTRRYINEAGLANYVRILDTLLAASIQPVTIYHMDL 1496
QY 128 PQLAHRYGWLNVVEYQDLPERYARLCFERFGDRQVNTTIXPVIQAIYATGSNAP 187
Db 1497 PQLAQD-VGGWEN-ETIVORFKEYADVLFQRLDCKVFWITLNEPFIAYAGYGYTAAP 1554
QY 188 GRSISNGSTEGTATEPMLAGKAQIMSHARAAVAVSRDPSOKQIGISLNGDYEPW 247
Db 1555 GVSN-----RPGTAPYVIGHNLIKAAEAHNLINDVYRAQGVITITISDVAEPR 1606
QY 248 DSNEPRDKAEAEERMEFHIGWPAFPIFLKKDYPESMKKQIG-----RLPALTPA 297
Db 1607 DPSQEDVEAARRYVQFMGSMFAPIFKNGDVNEVMKTRIDRSLAAGLKSRLPERTES 1666
QY 298 DFLANAGETDPQMGVYYSQPARHLDGVPPEVDYLGAIHEHENDGSGVGE-----EGG 353
Db 1667 EKRRIN-GIYDFEFGFHHYTTVLAVNL-----NYATLASSFPADGVASIAQSRMPDGG 1718
QY 354 LAMLRSCPDWFRGRLRVYGLVYG-KPIYITENGCCPGEENMTCEAAVNDPFRIRFDSH 412
Db 1719 SFMLKMTFRGFRILMLWKEEYNDPIYITENG-----SCREBTDIADTAIYIYLRTY 1772
QY 413 LDSISKAITODGVVKGYPFAMALLDNLWSDGYGPRGVTFTDYT--TLKRTPKXA 467
Db 1773 INEALXAV-ODKVDLGRYVWSAMDNEFWATGFSERFGHFFVYSDPSLPRIIPKASA 1828

RESULT 5
LPH_RABIT STANDARD; PRT; 1926 AA.
ID LPH_RABIT
AC P09849;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Lactase-phlorizin hydrolase precursor (Lactase-glycosylceramidase)
DE [includes: Lactase (EC 3.2.1.108); Phlorizin hydrolase (EC 3.2.1.62)].
GN LCT OR LPH.
OS Oryzococcus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryzococcus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=New Zealand white.
RX MEDLINE=89030634; PubMed=2460343;
RA Mantel N., Villa M., Enzler T., Wacker H., Bohl W., James P.,
Hunziker W., Semenza G.;


```

CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL: X56748; CAA40070.1; -
CC EMBL: X56747; CAA40069.1; -
CC EMBL: L04635; AAA41539.1; -
CC PIR: J05610; J05610.
CC HSSP: P26205; 1CBG.
CC InterPro: IPR001360; Glyco_hydro_1.
CC Pfam: PF00232; Glyco_hydro_1; 5.
CC PRINTS: PR00131; GLYDRLAS1.
CC ProDom: PD000650; Glyco_hydro_1; 4.
CC PROSITE: PS00572; GLYCOSYL_HYDROL_F1_1; 1.
CC PROSITE: PS00653; GLYCOSYL_HYDROL_F1_2; 2.
CC HydroLase; Glycosidase; Zymogen; Signal; Transmembrane; Repeat.
KV SIGNAL 1 19
KV PROPEP 20 867 BETA-GLUCOSIDASE.
FT CHAIN 868 1928 LACTASE-PHOSPHATIDYL HYDROLASE.
FT DOMAIN 22 1883 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1884 1902 POTENTIAL.
FT DOMAIN 1903 1928 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 89 1842 4 X APPROXIMATE REPEATS.
FT REPEAT 89 175 1.
FT REPEAT 365 849 2.
FT REPEAT 886 1366 3.
FT REPEAT 1371 1842 4.
FT ACT_SITE 1067 1067 PROTON DONOR (POTENTIAL).
FT ACT_SITE 1274 1274 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 1539 1539 PROTON DONOR (POTENTIAL).
FT ACT_SITE 1750 1750 NUCLEOPHILE (BY SIMILARITY).
FT CONFLICT 7 7 A -> R (IN REF. 1; AAA41539).
FT CONFLICT 113 113 Q -> E (IN REF. 1; AAA41539).
FT CONFLICT 207 207 E -> D (IN REF. 1; AAA41539).
SQ SEQUENCE 1928 AA; 217266 MW; 56DDCAAC44CAE85 CRC64;

Query Match 32.2%; Score 851.5; DB 1; Length 1928;
Best Local Similarity 40.5%; Pred. No. 8.9e-58;
Matches 193; Conservative 71; Mismatches 170; Indels 43; Gaps 13;

QY 8 PNDEWGGATAAYOIEGAVKEGGRPSIMDTYCHLEPERTGANGDVACDHYHRYDEPDFD 67
Db 1379 PKGFIWASAASAYVEGAMRADGKLSIMDTFSRPLRIGDNDGDVACDSYHKAIEDV 1438
QY 68 LITRYGAKAYRPSLSWSRIIPDLGRLDNEEGIEFYSKLIDALLRIGITPWTLYHMDL 127
Db 1439 ALQWIGVSHYRFSIAMSRIIP-DGTTKFINAGLSYVRFIDALLAGITPQVTIYHMDL 1497
QY 128 POALHRRGGLNVEVOLDERRARLCFERPGGRVQWITIKXPMIOAITGYATGSNAP 187
Db 1498 POALQD-VGWMEN-ETIVQREKAYADVLFQRLGDRVKKWITLNEPVIIAAGYGTGSAP 1555
QY 188 GRSSINKSTEGNTATEPMLAKQAQIMSHARAVAVSRDPSPQKQIGISLNGDYEPW 247
Db 1556 GISF-----RPTATYIAGHNLIKHAKHAKHLYNDVYARQCGTISITSSDWGEPR 1607
QY 248 DSNBPRKEAERMEFHIGWPNPIFKQDPESMK-----KQLG-----ERLPALTPA 297
Db 1608 DPTREHREARASYQVFMGFAHPIFKNGDYPEVMKTRIRDRISGLAKLNSRLPEFTES 1667
QY 298 DPAIIMGETDYGWNYTSGFARHLDPVETTYLGIHGHQNKQSPGEE-----SG 353
Db 1668 EKSRIR-GRFDFPGFNHNTVLAVNL-----DYPAFSSFDADRGVAIADSSWPVSG 1719
QY 354 LAMRSCPDMPERKLARVYGLYK-PIYTENGCPCEBEMWTCBEAVNDPFRIRYFDSH 412
Db 1720 SFWLKVPDFGRRIILNMLKEEYNNPIYVTENGSRREPEL-----NDIDRIYLRSY 1773
QY 413 LDSISKAITQDGVVVKGYFAMALLDNLEWSDGYGRGCVTPTDYT--TLKRTPKKSA 467
Db 1774 INEALKAV-HDKVLDLGRYTVMSIMDNFEMATGFAERFCVHVNNSDPELPRIPRASA 1829

```

```

ID MYO_ARATH STANDARD; PRT; 541 AA.
AC P37702; O987H2; Q93231; Q940N8;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosinase precursor (EC 3.2.1.147) (sinigrinase) (thiolglucosidase).
GN TCGL OR AT5G26000 OR T1N24.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
EX MEDLINE=94302158; PubMed=8029343;
RA Chadehwan S., Bishop J., Thangstad O.P., Bones A.M.,
RA Mitchell-Olds T., Bradley D.;
RT "Arabidopsis cDNA sequence encoding myosinase.";
RL Plant Physiol. 103:671-671 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=leaf;
RX MEDLINE=95284348; PubMed=7766881;
RA Xue J., Joergensen M., Philgren U., Raak L.;
RT "The myosinase gene family in Arabidopsis thaliana: gene
organization, expression and evolution.";
RL Plant Mol. Biol. 27:911-922 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Korani H., Kato T., Asamizu B.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Natsu K., Okumura S., Shino S., Takeuchi C., Wada T.,
RA Wakabayashi A., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston J.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneman T., Pegin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cortes M., Courtney L., Courtney L., Darce M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Stromatt C.,
RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
RA Mattiensen R., McCombie W.R., Wilson R.K., Murphy G., Banerjee I.,
RA Volckaert G., Wambolt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Enliken K.-D., Terry N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Ransberger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staveren M., Dirke W., Moollman P., Klein lankhorst R.,
RA Weltzenegger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
RA Feldpusch W., Lamberth S., Villarroel R., Gieles J., Ardiles W.,
RA Berts O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
thaliana.";
RL Nature 408:823-826 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onders C.S., Deng J.H., Akiyama K., Ansari Y.,
RA Akawa T., Bani J., Banno F., Bowser L., Brooks S.Y., Carrinci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jhang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Natsuka M., Seki M., Sekurai T.,

```

RESULT 7
MIRO_ARATH


```

RA  Satou M., Tanabe R., Vaysberg M., Wallender E.K., Mong C., Yamamura Y.,
RA  Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,
RT  "Empirical analysis of transcriptional activity in the Arabidopsis
RT  genome." ;
RL  Science 302:842-846(2003) .
RN  [5]
RN  RP  SEQUENCE OF 1-226 FROM N.A.
RA  Stracke R., Palme K. ;
RT  "Signal peptide selection derived cDNAs from Arabidopsis thaliana
RT  leaves and guard cells." ;
RL  Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN  [6]
RN  RP  SEQUENCE OF 192-250 FROM N.A.
RA  STRAIN=cv. C24; TISSUE=flower buds;
RA  Mache R., Quigley F., Thomas F., Yu D.Y. ;
CC  Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: Degradation of glucosinolates (glucose residue linked by
CC  a thioglucoside bound to an amino acid derivative) to glucose,
CC  sulfate and any of the products: thiocyanates, isothiocyanates,
CC  nitriles, epithionitriles or oxazolidine-2-thiones.
CC  -1- CATALYTIC ACTIVITY: H(2)O + a thioglucoside = a sugar + a thiol.
CC  -1- SUBUNIT: Homodimer (by similarity) .
CC  -1- SUBCELLULAR LOCATION: Vacuolar (by similarity) .
CC  -1- ALTERNATIVE PRODUCTS:
CC  Event=Alternative splicing; Named isoforms=1;
CC  Comment=A number of isoforms are produced. According to EST
CC  sequences;
CC  Name=1;
CC  IsoId=AF37702.1; Sequence=Displayed;
CC  -1- SIMILARITY: Belongs to family 1 of glucosyl hydrolases.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/isb-sib.ch).
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, L11454; AAC18869.1; -
DR  EMBL, X79194; CAAS5786.1; -
DR  EMBL, AF149413; AAD40143.1; -
DR  EMBL, AY045681; AAK74039.1; -
DR  EMBL, AY054237; AAL06896.1; -
DR  EMBL, AY058182; AAL25596.1; -
DR  EMBL, AY090382; AAL91284.1; -
DR  EMBL, AF083677; AAN60236.1; -
DR  EMBL, Z18232; CAAT9143.1; -
DR  PIR, S56653; S56653.
DR  HSSP, P26205; 1CBG3.
DR  SWISS-2DPAGE, P37702; ARATH.
DR  InterPro, IPR001360; Glyco_hydro_1.
DR  Pfam, PF00232; Glyco_hydro_1.1.
DR  PRINTS, PRO0131; GLHYDRLASE1.
DR  ProDom, PD000650; Glyco_hydro_1.1.
DR  PROSITE, PS00572; GLYCOSYL_HYDROL_F1.1; 1.
DR  PROSITE, PS00653; GLYCOSYL_HYDROL_F1.2; 1.
KW  Hydrolase; Glycosidase; Signal; Glycoprotein; Alternative splicing.
FT  SIGNAL 1 19
FT  CHAIN 20 541
FT  ACT_SITE 420 420
FT  CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT  CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT  CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT  CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT  CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT  CARBOHYD 379 379 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT  CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT  CARBOHYD 512 512 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT  CONFLICT 385 385 P -> H (IN REF. 4; AAL25596) .
FT  CONFLICT 426 426 P -> A (IN REF. 4; AAL25596) .
SQ  SEQUENCE 541 AA; 61132 MW; 3736B735DE7A5BD1 CRC64;

```

Query Match 32.1%; Score 849.5; DB 1; Length 541;
Best Local Similarity 39.5%; Pred. No. 2.4e-58;
Matches 185; Conservative 87; Mismatches 157; Indels 39; Gaps 14;

OY 11 FEMGFATAVAOIEGAVKEGGKPSIMPTCYHLEPSTNGA---NGDVAChYHRNDEDF 67
Db 47 FIFGVASSAVQVEG-----RGLGNVWDSFHHRPFEK-CGADLIGNDTTCOSYLTMQDID 102

OY 68 LLTYYGAKAYRFESLSMRRIIPLGRLDPVNEEGIEFYSKLIDALLRGITPMWTLYMWDL 127
Db 103 VMDELINSTGRFFSIAWSRLLPKGRSRGVNPGAIKYXNGLIDGLVAAKNMPFTVLFFMWDL 162

OY 128 POALHDHYGYGLANLEYQLDPEERARLCFPRFGRGVONMTINKPWIALIYGATGSNAP 187
Db 163 PQTLDENGENGLNKTTIVD-DPKDADACCFELFGVRVKNMHTTINOLYVPTRIGVALGTAP 221

OY 188 GRSS--INKSTESTNTATEPLMLAGKAQIMSHARAVAYSDFRPSQKGQIGISLNGDYE 245
Db 222 GRCPKPIDVRCPGNSSTTEPIYAHNOILLMAAADVRRRTKYKDQGMIQGPVITRMFL 281

OY 246 PWDSNEPPDKAEAREMEFHIGFANDIFLKQYPESMKCOLGERLPALTPADPAIIINAG 305
Db 282 PFDSHQ-EISKATERRAKIFFHGWMCELTGECK-YPDIMREVGCRDLREFSETEALVK-G 338

OY 306 ETDYGYNNVTYSQPARHLIDGVEPTDYLGAIHET---QENKOSPVE-----ES 352
Db 339 SYDLPLGNVYTQYANQNQTIVPSDVHTALMDSTRLLTSKNAATHADGPPNNAAYPYK 398

OY 353 GLAMLRSCPDPRFGHLARVGLYKRP-IYITENGCCPGSENNMCCEAVNDPFIRFYDFS 411
Db 399 GIYYVM---DYFKT-----TYDDPLLIYTENGFSPGDGF--EKATVDYKRIDLCS 446

OY 412 HLDISIKAITQDGVVVKYFPAMALLDNLEMSDGYGPRFGVTEDYTL 459
Db 447 HLCLFSKVYEKENVNKVFAMSLGDYVEFCNGFTVFFGLSYVDPAN 494

RESULT 8
MYR3_SINAL
ID MYR3_SINAL STANDARD; PRY; 544 AA.
AC P23092;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myrosinase MB3 precursor (EC 3.2.1.147) (Sinigrinase)
OS (ThioglucoSIDase).
OS Sinapis alba (White mustard) (Brassicaceae).
OC Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosoids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=cv. Maxi; TISSUE=Seed.
RX MEDLINE=92119259; PubMed=1731996;
RA Xue J.; Lemmen M.; Falk A.; Raak L.;
RT "The glucosinolate-degrading enzyme myrosinase in Brassicaceae is
RT encoded by a gene family.";
RU Plant Mol. Biol. 18:387-398(1992).
CC - FUNCTION: Degradation of glucosinolates (glucose residue linked by
CC a thioglucoside bound to an amino acid derivative) to glucose,
CC sulfate and any of the products: thiocyanates, isothiocyanates,
CC nitriles, allylthionitriles or oxazolidine-2-thiones.
CC - CATALYTIC ACTIVITY: H(2)O + a thioglucoside = a sugar + a thiol.
CC - SUBUNIT: Homodimer.
CC - SUBCELLULAR LOCATION: Vacuolar.
CC - TISSUE SPECIFICITY: In vacuoles called myrosin grains of a certain
CC class of cells, myrosin cells, distributed in the cotyledons and
CC the axis of the embryo as well as in different organs of the
CC growing plant.
CC - SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC This SWISS-PROT entry is copyright It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).

DR EMBL; X59879; CAA42534.1; -.
DR PIR; S19149; S19149.
DR HSP; P26205; ICBG.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F2_1; 1.
KM Hydrolyase; Glycosidase; Multigene family; Glycoprotein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 544 MYROSINASE MB3.
FT ACT SITE 426 446 NUCLEOPHILE (BY SIMILARITY).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 544 AA; 62050 MW; 3971D304CF75C28 CRC64;

Query March 32.0%; Score 845; DB 1; Length 544;
Best Local Similarity 39.3%; Pred. No. 5.3e-58;
Matches 190; Conservative 86; Mismatches 158; Indels 50; Gaps 15;

QY 10 DFEWGFATTAAYQIGAVVEGGRPSIMDTYCHLESRRING--ANGDVACDHYHREDFD 67
DB 48 DFIGVASSAQIIEG---RGRGVNVMDGFSHRPEKSGSLKNGDTSCESTYMKDVE 104
QY 68 LITRYGAKAYRFSLSWRIIPLGGRDLPVNEEGIEFYSKIDALLRGTTPWTLVHMDL 127
DB 105 IMGELNATGYFSPFAMSRIVPKGKVRGVDQAGLDYHNLDALLEKNITPFTVLFHMDL 164
QY 128 PQLAHADYVGWLVNEEVDLFEERYALCFERFGRDVRQWMTITNKPWIAIYGYATGSNAP 187
DB 165 PQLQDEVEGFLDKQIIQ-DKDYADLCFEKFGKVKMTITNQLYTVPTRGVAGTDAP 223
QY 188 GRSSINKHSTE---GNTATEPMLAKGAQIMSHARAANVAVSRDPRPSQKQIGISLNGDY 243
DB 224 GRGCPXYDTKRCRCGNSSTERYIAHQQLAHAIYDLVYTNV-AFQNGKIGVMITRW 282
QY 244 YEPWDSNEPRDKAERRMERFHIQMFANPIFLKKDYPSMKKQIGERLPALTPADFAILN 303
DB 283 FLPYDESDPACIEAERNQCFHGMWMEPL-TKGRYPDIMEQIVGSRLPNFTTEAF-AELV 340
QY 304 AGEITDFGMNYYTSGFAHLDGPVP-ET-----DVLGAIHHEQEKD 344
DB 341 AGSYDFGLNLYVYQAKPKNPYPSETHALMDAGVDLTNNRSGEGPGEVFAEDANSY 400
QY 345 GSPVGEESGLMLNRCSPDMFRKHLARVYGLYKGP-IYITENGCCPGGEENNTCEAVNDP 403
DB 401 YYPKG-----IYVMDYFKTK-----YNNPLIYITENGISTPSES-RC-EALADY 444
QY 404 FRIYFDSHLDSISKAITQDGVNVKGFAMALLDNLEWSDGYGRPFVGTFTDYTTL-KRT 462
DB 445 KRINVLCSHLCFLRKVIREKGVNIRGFYAMALGDNVECFCKGTVRFGLSYVNWMDLDRN 504
QY 463 PKKS 466
DB 505 LKES 508

RESULT 9
BGLA_THEMA STANDARD; PRT; 446 AA.
AC Q08638;
DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Beta-glucosidase A (EC 3.2.1.21) (Gentlobiase) (Celllobiase) (Beta-D-
DE glucoside glucosylhydrolase).
GN BGLA.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=94104595; Pubmed=8277941;
RA Liebl W., Gabelsberger J., Schleifer K.H.,
RT "Comparative amino acid sequence analysis of Thermotoga maritima
RT beta-glucosidase (BglA) deduced from the nucleotide sequence of the
RT gene indicates distant relationship between beta-glucosidases of the
RT BGLA family and other families of beta-1,4-glycosyl hydrolases.";
RL Mol. Gen. Genet. 242:111-115(1994).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC -1- glucose residues with release of beta-D-glucose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- MISCELLANEOUS: THE DNA CODING FOR THIS PROTEIN IS NOT FOUND IN THE
CC COMPLETE GENOME OF T.MARITIMA. IT COULD HAVE ORIGINATED FROM
CC ANOTHER BACTERIAL SPECIES.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).

DR EMBL; X74163; CAA52276.1; -.
DR PIR; S41561; S34570.
DR HSP; P22073; ITR1.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F2_1; 1.
KM Hydrolyase; Glycosidase; Cellulose degradation.
FT ACT SITE 166 166 PROTON DONOR (POTENTIAL).
FT ACT SITE 351 351 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 446 AA; 51548 MW; 2E5B06E72B84C01 CRC64;

Query March 31.7%; Score 839; DB 1; Length 446;
Best Local Similarity 41.1%; Pred. No. 1.2e-57;
Matches 193; Conservative 76; Mismatches 145; Indels 56; Gaps 19;

QY 8 PNDFWGFATTAAYQIGAVVEGGRPSIMDTYCHLESRRINGANGDVACDHYHREDFD 67
DB 7 PEGFLMGVATAAYQIESSPLADGAGMSIMWTFSTPNVNGKGDGDAVACHYNNWKEIDIE 66
QY 68 LITRYGAKAYRFSLSWRIIPLG-GRDLPVNEEGIEFYSKIDALLRGTTPWTLVHMD 126
DB 67 IIEKLGAKAYRFSLSWRIIPLEGTR---VNGKGLDPTNNKIIDTLLEKGIPTFTIYHMD 123
QY 127 LPOLAHADYVGWLVNEEVDLFEERYALCFERFGRDVRQWMTITNKPWIAIYGYATGSNA 186
DB 124 LPPALQK-GGMAN-REIADWFAEYSRVLFENFDKRVKMTITNEPWWVAIVGHLVGVHA 181
QY 187 PGRSSINKHSTEGNTATEPMLAGKA---QMSHARAANVAVSRDPRPSQK-GQIGISLNGD 242
DB 182 PGMKDI-----YVAFRAVHNLRAHRAVAV-----FRETVMKQKGIIVENNG 224
QY 243 YIEPWSNEPRDKAERRMERF-HIGMFANPIFLKKDYPSMKKQIGERLPALTPADFAI 301
DB 225 YIEP-ASEKSEDIRAVAFHGFQNNYPLFLNPIY-RGIYPELVLEFAAEYLPENYKDDMS 282
QY 302 LNAGETDFGMNYYTSGFAHLDGPVPETDYLGAIHHEQEKDGPVGE-----SGLAMLR 358

Db 283 IQE-KIFVGLINYSGLVLF-----DPPAPAKYSEVERLPTKAMGWEIVPEGIYWI- 334
Qy 359 SCPMRFKHLLARVYGLYKGP-IYITENGCPGCGENMTCEBA-VNDFRIYRFPMSHDSI 416
Db 335 -----LKKVEEENPEVYITENGA---AFDVADEGRRHDNDYLDKHAHQGA 382
Qy 417 SKAITQGVVVKGYFAMALLDNLEMSDYGPRFVTFDYTTLLKRTPKKS 466
Db 383 WKAI-QGCVPLKGFVMSLDNPFEMAGYSKRFIVVVDYSTQRIYKDS 431
RESULT 10
BGL_1 ARATH STANDARD; PRT; 528 AA.
AC Q3SE50;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-glucosidase homolog precursor (EC 3.2.1.-).
GN BGI OR ARI52400 OR F19K6.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Pao H.L., Hwang I.;
RT "A transgenic Arabidopsis plant overexpressing an ER localized b-
RT glucosidase homolog that is transcriptionally suppressed by NaCl is
RT hypersensitive to NaCl stress."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA STRAIN=cv. Columbia;
RC MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Becker J.R., Palm C.J., Federpiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etzu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gail J.E., Geldmuth A.D., Haas B., Hansen N.F., Hughes B., Hutzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malti R., Marzilli A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.U., Tambunga G., Toriumi M.J., Town C.D.,
RA Uteback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum.
CC -1- INDUCTION: Transcription suppressed by NaCl.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF183827; AAF22295.1; -
DR EMBL: AC037424; AAG51546.1; -
DR HSSP: P26205; ICBG.
DR SWISS-2DPAGE: Q3SE50; ARATH.
DR InterPro: IPR000886; ER_target_S.

DR InterPro: IPR001360; Glyco_hydro_1.
DR Pfam: PF00232; Glyco_hydro_1.
DR PRINTS: PR00131; GLYDRASE1.
DR ProDom: PD000650; Glyco_hydro_1.
DR PROSITE: PS00014; ER_TARGET_1.
DR PROSITE: PS00572; GLYCOSYL_HYDROL_F1_1; FALSE_NEG.
DR PROSITE: PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KM Hydrolyase; Glycosidase; Endoplasmic reticulum; Glycoprotein; Signal.
FT SIGNAL 1 21
FT CHAIN 22 528
FT SITE 525 528
FT ACT_SITE 207 207
FT ACT_SITE 422 422
FT CARBOHYD 189 189
FT CARBOHYD 466 466
FT CARBOHYD 499 499
FT CONFLICT 134 134
SQ SEQUENCE 528 AA; 60486 MW; 43871070F66C1140 CRC64;
Query Match 31.4%; Score 831; DB 1; Length 528;
Best local Similarity 39.4%; Pred. No. 6, 2e-57;
Matches 191; Conservative 67; Mismatches 185; Indels 42; Gaps 14;
Qy 5 LALPDPFEMGFATTAAYQIEGAVKEGGRPSIMDYCHLEBSRTNGANGVACDHYHRDE 64
Db 42 LNFPEGFIMGTATAAQVEGAVNEGCRGPMWDTFKKPFHRCENHNADVAVDFYHRYKE 101
Qy 65 DFDLTKYGAAYRSLMSRRIPLGGRDPVNEBIEPYSKILDLRRGTPWTLVH 124
Db 102 DIDLMDLNTDARLSIAMPRIIPHRMSKGIKVGQFYHDLIDELKNIIPLVTVFH 161
Qy 125 WDLPOLHDEYGGMLNVEEYQDPEERYARCFERFGRVONMTITXNPWIAQIYVATGS 184
Db 162 WDFPOLDEYGGFLSGRIYO-DFTYANTFHEHGKVMITFNEPWFSAIDNGK 220
Qy 185 NAFGRSS-----INKSTEGNTATEPMLAGKAQIMSHARAAYASDFRPSQKQIGISL 239
Db 221 KARGRSPYIPGVGQCDQDRSGEYAVQVSHNLLSHAVVADF-RNCKQACAGKIGIAH 279
Qy 240 NGDIYEPWMSNERDEAERMEFHIGFANFIPLKDYPSMKQLGERLPAALPADF 299
Db 280 SPWFPEFQDLEHVG--SIERVIDFLIGMLAFT-TYGDVPSMKRVGRRLKFEAEK 336
Qy 300 AINAGETDPFGNNYTSOPARHLDDGPVE-----TDVGAHHEQNKGSVGE-- 351
Db 337 KLKLK-GSTDVGNVNYTSVFAKEI-SPDPKSPWTTDSL-VWDKSKVDGYKIGKFPN 392
Qy 352 -----SGIAMPSCDPMFRKHLARVYGLYKGP-IYITENGCPGCGENMTCEAVND 402
Db 393 GKLDVYSKGIKRYL-----LKYIKDNYGDPVEVITAEENGGEDELGEKHNVDFTOD 442
Qy 403 PPIRYPDSHDSISKAITQGVVVKGYFAMALLDNLEMSDYGPRFVTFDY-TTLKR 461
Db 443 HNRKYVIGRHLISMHALCKDKVNVGVFWMSLDNPFEMQDGYKARFGLYIDFQNNLTR 502
Qy 462 TPCKS 466
Db 503 HQKVS 507
RESULT 11
BGL_1 ARATH STANDARD; PRT; 444 AA.
AC Q33843; O52505;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-glucosidase A (EC 3.2.1.21) (gentiobiose) (Cellulobiose) (Beta-D-
DE glucoside gluconylolase).
GN BGLA OR GGAH.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_Taxid=2337;

Db 124 POALQDQ-GCM--GSRITIDAFEAELMEKELGKIKOMITTEPWCMAFLSNYLGVHA 180
 Qy 187 PGS-----SINKHSTEGTATEPMLAGKAQIMSHARAVASRDPRPSOKOIGISLN 240
 Db 181 PGKNDQLAIDVSHH-----LVAHGRAVTLF-RELGIS--GEIGIAPN 221
 Qy 241 GDYVEPWSNPPDKAEARMEFHIGWFPANPIFLKKDYPSM--KKOGERIPALTPA 297
 Db 222 TSMAVPRRTK-EDMECCLRVNGMSGWYLDPIYF-GEYKPFMDWENIGYK-PRIVDG 278
 Qy 298 DPAIILNAGETDPYGMVNYTSQFARHLDPVPETDYLGAIHEHOENKDGSPVGEESGLAWL 357
 Db 279 DMLILG-PIDFICINYYTSSMKNYNPGEA-----CGMLSSEASISWGAP--KTIDIGM- 327
 Qy 358 RSCPEMRKRLARYGL-----YGRP-IYITENGCPGCEENMTCEEAIVNDPRIRYF 409
 Db 328 ----EYAEGL--YDLRLYADKRYGNPTLYITENGA-C-YNDGLSLDGRHIDRRIDYL 378
 Qy 410 DSHLDSISKAITODGVVVKGYFAMALLDNLEMSDGYPRFGVTFTDTTLKRTPKKS 466
 Db 379 AMHLIQASRAI-EDGINLKGKYMESLMDNFEMAGCYMRGLVHVVDYDILVRTPKDS 434

RESULT 13

BGLB_PAEPO STANDARD; PRT; 448 AA.

ID BGLB_PAEPO
 AC P2505;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE Beta-glucosidase B (EC 3.2.1.21) (Gentloblase) (Cellloblase) (Beta-D-glucoside glucosylhydrolase) (Amygdalase).
 GN BGLB.
 OS Paenibacillus polymyxa (Bacillus polymyxa).
 OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
 OX NCBI_TaxID=1406;
 RN [1]
 RP MEDLINE=91071603; Pubmed=2123813;
 RA Gonzalez-Candelas L., Ramon D., Polaina J.;
 RT "Sequences and homology analysis of two genes encoding beta-glucosidases from Bacillus polymyxa."
 RL Gene 95:31-38(1990).
 CC -1- CATALYTIC ACTIVITY: Hydrolase of terminal, non-reducing beta-D-glucose residues with release of beta-D-glucose.
 CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC EMBL: M60211; AAA23264.1; -
 DR PIR; J00038; J00038.
 DR HSSP; P22073; ITRI.
 DR InterPro; IPR001360; Glyco_hydro_1.
 DR Pfam; PF00232; Glyco_hydro_1; 1.
 DR PRINTS; PR00131; GLHYDRLASE1.
 DR Prodom; PD000650; Glyco_hydro_1; 1.
 DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
 DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
 KW Hydrolase; Glycosidase; Cellulose degradation.
 FT ACT_SITE 167 PROTON DONOR (POTENTIAL).
 FT ACT_SITE 356 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 448 AA; 51573 MW; 628F44507C21EF1F CRC64;

Query Match 29.9%; Score 791; DB 1; Length 448;
 Best Local Similarity 38.1%; Pred. No. 6.4e-54;
 Matches 185; Conservative 61; Mismatches 172; Indels 68; Gaps 15;

Qy 8 PNDWFGFATAAYQIEGAVEGGRSPINDTYCHLEPSRTNGANGVACDHYHRVDEDFD 67
 Db 9 PATFMGTSTSSQIIGGDEGRTSPINDTFQIQKVLGCGGVAQCHPHFKREDVQ 68
 Qy 68 LITKYGAKAYRFLSRSRIIPLGRLDPVNEEGIEFYSKLIDLLRGRITPMVTLTHMDL 127
 Db 69 LMKQFLFHLHFRFSVAMPRIIMPAG---IINEGLFVEHLLDEIELAGLPMULTVHMDL 125
 Qy 128 POALHRYGGMVAVEVQDFFERYARLCPEFGDRVQNNITTXPMIOAIYVATGSMA 187
 Db 126 PQMTEDB-GGWTORETIQ-HFKTYASVIMDRFERRINMNTINBPYCASLIGTGEHAP 183
 Qy 188 GRSINKHSTEGTATEPMLAGKAQIMSHARAVASRDPRPSOKOIGISLNGDYEPW 247
 Db 184 GHENMR-----EAFPAHHLICMGIAANLHK---EKGLTGKIGTLMEHAYDA- 229
 Qy 248 DSNPPDKAEARMEFHIGWFPANPIFLKKDYPSMKKOLGERLPAL--TPADFAILNA 304
 Db 230 ASRRPEVAAAIRRDGFINRWFAEPLFNCR-YPEDWVEWYGYLNGIDFVQPDMLIQ 288
 Qy 305 GEDTFGMNYTSQFAR-----HLDGVPETDYLGAIHEHOENKDGSPVGE 351
 Db 289 -PDGFLGINYYTSITRSTNDASLQVEQVHMEEPV--TDMGWEIH----- 331
 Qy 352 SGLAWLRSCPEMRKRLARYGLYGR--PIYITENGCPGCEENMTCEEAIVNDPRIRYF 409
 Db 332 -----PESTFKLITREKDFSGKGLPILITENGAARDE---LVNGQIEDTGRHGYI 379
 Qy 410 DSHLDSISKAITODGVVVKGYFAMALLDNLEMSDGYPRFGVTFTDTTLKRTPKSALV 469
 Db 380 EEHLKACHREI-BEGQLKGYFWMSFLDNFEMAGYSKRGIVHINVEOTERTPKOSALW 438
 Qy 470 LKDMFA 475
 Db 439 FKQWMA 444

RESULT 14

BGLT_TRIRP STANDARD; PRT; 425 AA.

ID BGLT_TRIRP
 AC P26205;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cyanogenic beta-glucosidase precursor (EC 3.2.1.21) (Linamarase) (Fragment).
 GN LT.
 OS Trifolium repens (Creeping white clover).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC Eurosidia I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Trifolium.
 OX NCBI_TaxID=3899;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 12-25 AND 125-147.
 RC STRAIN=5100 (EG); TISSUE=leaf;
 RX MEDLINE=91322517; Pubmed=1907511;
 RA Oxtoby E., Dunn M.A., Pancoro A., Hughes M.A.;
 RT "Nucleotide and derived amino acid sequence of the cyanogenic beta-glucosidase (linamarase) from white clover (Trifolium repens L.).";
 RL Plant Mol. Biol. 17:209-219 (1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
 RX MEDLINE=96097401; Pubmed=8535788;
 RA Barrett T., Suresh C.G., Tolley S.P., Dodson E.J., Hughes M.A.;
 RT "The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase.";
 RL Structure 3:951-960 (1995).
 CC -1- FUNCTION: Hydrolases cyanoglucosides, contributing to the release of hydrocyanic acid, which functions as a defense mechanism against small predators, when the leaf tissue is damaged.
 CC -1- CATALYTIC ACTIVITY: Hydrolase of terminal, non-reducing beta-D-glucose residues with release of beta-D-glucose.
 CC -1- PATHWAY: Cyanogenesis.

CC -1- SUBUNIT: Homodimer.
CC -1- TISSUE SPECIFICITY: Leaves.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56733; CAA40057.1; -
DR PIR; S16580; GLJY14.
DR PDB; 1CBG; 15-OCT-95.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLYDRLASE1.
DR PRODOM; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1
FT NON TER 1
FT CHAIN 1
FT ACT_SITE 12 425
FT ACT_SITE 194 425
FT ACT_SITE 408 408
FT DISULFID 213 221
FT CARBOHYD 220 220
FT CARBOHYD 412 412
FT TURND 16 17
FT HELIX 22 24
FT HELIX 27 29
FT TURND 32 33
FT STRAND 35 39
FT HELIX 42 45
FT TURND 52 53
FT STRAND 57 57
FT HELIX 58 65
FT HELIX 67 69
FT TURND 71 72
FT TURND 79 80
FT HELIX 82 95
FT TURND 96 97
FT STRAND 100 104
FT HELIX 107 110
FT TURND 112 113
FT HELIX 116 118
FT HELIX 122 137
FT TURND 138 139
FT STRAND 141 147
FT STRAND 152 152
FT HELIX 153 159
FT HELIX 161 163
FT TURND 165 166
FT HELIX 167 182
FT TURND 183 185
FT STRAND 188 193
FT HELIX 195 203
FT TURND 204 204
FT TURND 209 210
FT HELIX 215 217
FT TURND 226 228
FT HELIX 252 251
FT TURND 252 252
FT HELIX 253 256
FT STRAND 259 265
FT STRAND 268 271
FT HELIX 276 289
FT TURND 290 290
FT HELIX 291 299
FT HELIX 304 310
FT HELIX 311 313

FT HELIX 319 325
FT TURND 326 327
FT STRAND 331 335
FT STRAND 339 344
FT TURND 349 350
FT HELIX 355 358
FT TURND 359 359
FT STRAND 361 364
FT STRAND 366 367
FT TURND 368 369
FT STRAND 370 371
FT STRAND 375 375
FT TURND 379 380
FT STRAND 382 382
FT TURND 385 385
FT HELIX 386 398
FT TURND 399 400
FT STRAND 404 408
FT STRAND 413 414
FT TURND 417 418
FT HELIX 421 425
SQ SEQUENCE 425 AA; 48324 MW; 20B983B65C47A678 CRC64;

Query Match 29.9%; Score 790; DB 1; Length 425;

Best Local Similarity 43.9%; Pred. No. 7.1e-54; Matches 169; Conservative 63; Mismatches 135; Indels 18; Gaps 10;

Qy 11 FEMGPTAAVQIEGAVKEGGRPSIWDYCHLEPSR-TNGANGDVACDHVHYHDEDFDL 69
Db 34 FVEGTASAPQYEGGAFAFEDGKPSIWDTFHKKYPEKIKORTGDAVIDEYHKKEDIGIM 93
Qy 70 TKYGAAYRPSLSMSRIIPGGRDLPVNEEGIEFYKLIIDALIRGITWVTLHYMDLPQ 129
Db 94 KDMNLDATRPSISWPEVLPGKISGVNREGINYNLNEVLANGMOPEYVTLFHWDPQ 153
Qy 130 ALHDRVGMVINEVOLDPEFYARLCEFRFGDVQWMTITNXPWQAIYGYATGSNAPGR 189
Db 154 ALDEYRGFLGRNIVD-DFRDYAEICFKFGRGVGMITLNEPKGSMNAYVGFAPGR 212
Qy 190 SS--INKHTEGNTATEPWLAKQAQIMSHARAVAYSRRDFPSQKQIGISLNGDYEPW 247
Db 213 CSWMLNCTGSGSGREPYLAHYQLAHAAARLYKTYQASQNGIIGITLVSHWFEF- 271
Qy 248 DSNRPDKAEAEERBEFHIGWFPANPIFLKQVPSKQQLGRPLPALTPADPAIINAGET 307
Db 272 ASKEKADVDAAKRGDPMFGWFMHPL-TKGRYPESKRYLVKRLKFTSESKEL-TGSF 329
Qy 308 DFYGNVYTSQPARHLD----GPVPEFDYL-GAIHEHQENKDGSPVGEESGLAWLRSCP 361
Db 330 DFLGLNVYSSYYAAKAPRIIPARPAIQDSLINATFEH---NGKPLGPMMASSWLCIYP 385
Qy 362 DMFRKHLARVGLYKGP-IYITENG 385
Db 386 QGIRKLLLYKNHYNPNVYITENG 410

RESULT 15
BGLA_PAEPO
ID BGLA_PAEPO STANDARD; PRT; 448 AA.
AC P22073;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Beta-glucosidase A (EC 3.2.1.21) (Gentioibase) (Cellobiase) (Beta-D-
DE glucoside glucohydrolase) (Amygdalase) (BGA).
GN BGLA.
OS Paenibacillus polymyxa (Bacillus polymyxa).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxId=1406;
RP [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91071603; PubMed=2123813;
RA Gonzalez-Candelas L., Ramon D., Polaina J.;

RT "Sequences and homology analysis of two genes encoding
RT beta-glucosidases from *Bacillus polymyxa*.";
RL Gene 95:31-38(1990).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RC STRAIN=ATCC 842;
RA MEDLINE=98139963; PubMed=9466926;
RA Sanz-Aparicio J., Hermoso J.A., Martinez-Ripoll M., Lequerica J.L.,
RA Polaina J.;
RT "Crystal structure of beta-glucosidase A from *Bacillus polymyxa*:
RT insights into the catalytic activity in family 1 glycosyl
RT hydrolases.";
RL J. Mol. Biol. 275:491-502(1998).
CC -1- FUNCTION: BG4 IS INTRACELLULAR AND CLEAVES CELLOBIOSE PROBABLY
CC THROUGH INORGANIC PHOSPHATE MEDIATED HYDROLYSIS.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC glucose residues with release of beta-D-glucose.
CC -1- SUBUNIT: Homooctamer.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; J06037; AAA22263.1; -.
DR PIR; J06037; J06037.
DR PDB; 1BGA; 15-APR-98.
DR PDB; 1BGQ; 27-MAY-98.
DR PDB; 1TR1; 20-APR-99.
DR PDB; 1E41; 05-JUL-01.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR PRODOM; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F2_1; 1.
DR Hydrolase; Glycosidase; Cellulose degradation; 3D-structure.
KW ACT SITE 166 166 PROTON DONOR (POTENTIAL).
FT ACT SITE 352 352 NUCLEOPHILE (BY SIMILARITY).
FT STRAND 3 5
FT TURN 8 9
FT STRAND 11 15
FT HELIX 18 21
FT TURN 25 26
FT HELIX 27 29
FT STRAND 33 33
FT HELIX 34 39
FT TURN 40 40
FT TURN 42 44
FT TURN 46 48
FT HELIX 54 55
FT TURN 57 71
FT HELIX 72 72
FT TURN 75 79
FT STRAND 82 85
FT HELIX 87 88
FT TURN 95 110
FT HELIX 111 112
FT TURN 114 120
FT STRAND 125 125
FT HELIX 126 130
FT TURN 131 132
FT HELIX 133 135
FT TURN 137 137
FT HELIX 138 153
FT TURN 154 154
FT HELIX 155 157
FT STRAND 160 165
FT HELIX 167 175

FT TURN 176 176
FT TURN 181 182
FT HELIX 187 210
FT TURN 211 212
FT TURN 216 220
FT STRAND 222 222
FT STRAND 226 228
FT HELIX 233 246
FT TURN 247 247
FT HELIX 248 256
FT HELIX 261 269
FT TURN 270 271
FT TURN 278 279
FT HELIX 280 284
FT TURN 285 285
FT STRAND 290 294
FT STRAND 299 303
FT TURN 305 306
FT TURN 308 311
FT STRAND 312 314
FT STRAND 321 321
FT TURN 323 324
FT STRAND 327 327
FT TURN 330 330
FT HELIX 331 340
FT HELIX 341 343
FT STRAND 348 352
FT STRAND 362 362
FT TURN 363 364
FT STRAND 365 365
FT HELIX 369 387
FT TURN 388 389
FT STRAND 392 397
FT STRAND 401 401
FT HELIX 406 411
FT STRAND 417 419
FT TURN 421 423
FT STRAND 426 428
FT HELIX 430 441
FT TURN 442 442
FT STRAND 443 445
SQ SEQUENCE 448 AA; 51649 MW; D971D2B61B6627C1 CRC64;

Query Match 29.8%; Score 789; DB 1; Length 448;
Best Local Similarity 39.8%; Pred. No. 9, 1e-54;
Matches 184; Conservative 65; Mismatches 173; Indels 40; Gaps 18;

QY 8 PNDPEWGFATPAAYQIEGAVKEGGGSPSIWDYCYCHLEPSRTNGANGDVACDHYHRDDEFD 67
DB 7 PDDFMWGTAATAAYQIEGAVGEDRGSLINDTFAHTEGKVFNGDNGVACDSYHRYEEDIR 66
QY 68 LRTTKGAKAYRFLSRSRIIPLAGRLDPVNEBSIEEYSKIDMLRGLTPWVTLVHMDL 127
DB 67 LMKELGIRTYRFPVSVPRIFFPND--GEVNOEGLDYHVRVVDLINDNGIEPCTLYHMDL 124
QY 128 PQLAHDRYGGWLVVEVQDLEFERYARLCFERFGDRVQNNITTXPMIOIYGVATGSNAP 187
DB 125 PQLAHDA-GGNGKRRITQA-FVQFAETMREFFGKIQHMTTFNEPNCIAFLSMGLVHAP 182
QY 188 GRSSINKHSTEGNTATEPMLAGKQIMSHARAVVYSRDFRP-SQKQIGISLNGDYEP 246
DB 183 GLTNL-----QTAID---VGHHLVAVGLSV---RRPRELGTSGQIGIAPNVSWAVP 228
QY 247 WDSNEPRDKEAARRRNEPHIGWPAANIPLKDYPESMKQOLGRRLPALTPADPAIINAGE 306
DB 229 YSTSE-EDKAAACARTISLSHSDWFLQPIY-QGSYPOFLVWMAEQGATVPIDQDMOIG 286
QY 307 -TDFYGANYTTSQFAHLDGPVETDYIGAIEHGOENKDGSPVGESEGLAWLSCDFMFR 365
DB 287 PIDMIGINYSMSVNRP---NPEAGFL-----OSEIINGLPV---TDIGMPVESGLV 335
QY 366 -KHLARVYGLYGPRIYITENGCPGPEENMTCEAVNDPFRIRYFSDHLSISKAITODG 424

Mon Mar 29 09:02:57 2004

us-10-026-140-2.rsp

Page 14

Db 336 VLHYLQKYG--NIDIYITNGA-CINDE--VNGKQVQDRKISIMQGHLYQVHRTI-HDQ 389

QY 425 VVKGYSFAMALLNLEMSDQYGRFVETDITTLKTRPKS 466

Db 390 LHYGYSMAWGLDNFEMARGYNMRFGMLHVDPRQVTRPKES 431

Search completed: March 25, 2004, 15:55:37
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2004, 15:53:27 ; Search time 21 Seconds

(without alignments)
2216.986 Million cell updates/sec

Title: US-10-026-140-2

Perfect score: 2644

Sequence: 1 MPESLALPMDFEWGFATPAAY.....KSALVLKDMFARQVRVAA 484

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78:*

2: PIR1:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1037	39.2	510	2	TK7.7 protein - A
2	1003	37.9	577	2	beta-glucosidase-1
3	999	37.8	577	2	probable beta-gluc
4	984.5	37.2	591	2	beta-glucosidase h
5	983	37.2	551	2	beta-glucosidase (
6	976	36.9	509	2	beta-glucosidase B
7	965	36.5	560	2	probable beta-gluc
8	964.5	36.5	562	2	flurostanol glycosi
9	960.5	36.3	534	2	beta-glucosidase-1
10	944.5	35.7	517	2	probable beta-gluc
11	943.5	35.7	549	2	beta-glucosidase (
12	932	35.2	574	2	beta-D-glucosidase
13	928	35.1	547	2	beta-glucosidase (
14	927	35.0	519	2	beta-glucosidase (
15	925.5	35.0	506	2	probable beta-gluc
16	922	34.9	527	2	beta-glucosidase h
17	914	34.6	514	2	beta-glucosidase-1
18	913.5	34.5	565	2	probable beta-gluc
19	902	34.1	563	2	beta-D-glucosidase
20	896	33.9	531	2	beta-glucosidase (
21	884	33.4	566	2	beta-glucosidase (
22	880.5	33.3	493	1	beta-glucosidase (
23	877	33.2	524	2	probable beta-gluc
24	873	33.0	1920	2	lactase (EC 3.2.1.
25	868	32.8	524	2	probable beta-gluc
26	863.5	32.7	524	2	chitoglucosidase (E
27	863.5	32.7	1918	2	lactase (EC 3.2.1.
28	859	32.5	520	2	beta-glucosidase h
29	858.5	32.5	548	2	chitoglucosidase (E

30	857.5	32.4	1927	2	beta-glycosidase c
31	854	32.3	536	2	chitoglucosidase (E
32	852.5	32.2	1926	2	beta-glucosidase c
33	851.5	32.2	1928	2	beta-galactosidase
34	849.5	32.1	534	2	chitoglucosidase (E
35	845	32.0	544	2	chitoglucosidase (E
36	843	31.9	531	2	probable beta-gluc
37	839	31.7	446	2	beta-glucosidase (
38	831.5	31.4	544	2	chitoglucosidase (E
39	831.5	31.4	547	2	chitoglucosidase (E
40	826	31.2	528	2	probable beta-gluc
41	825	31.2	575	2	beta-D-glucosidase
42	825	31.2	882	2	protein T4012.15 l
43	819	31.0	535	2	probable beta-gluc
44	814	30.8	496	2	hypothetical prote
45	813.5	30.8	447	2	beta-glucosidase b

ALIGNMENTS

RESULT 1

TK7.7 protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001

C/Accession: F86392

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzar, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon, I

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: F86392

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-510 <STO>

A/Cross-references: GB:AE005112; NID:9797746; PIDN:AAF98564.1; GSPDB:GN00141

C/Genetics:

A:Map position: 1

C:Superfamily: Agrobacterium beta-glucosidase

Query Match 39.2%; Score 1037; DB 2; Length 510;

Best Local Similarity 44.1%; Pred. No. 4.3e-74;

Matches 209; Conservative 85; Mismatches 156; Indels 24; Gaps 10;

QY	6	ALPNDPEWGFATPAAYIQEGAVKGGKGRGPSITWDYCHLEPRTGANGDVAACHYHRDED	65
DB	36	SPFGKGFVFGTASAPFOHEGAVKAEGRGPTIMDFSTFGKITPFSNADVAVDDYHRRFED	95
QY	66	FDLLTKYGAAYRPSLSMSRIIPGKRLDPVNEGIEFYSKLDALLRSGITPMVTLTYH	125
DB	96	VQLMKMGMDAYFISWTRIFPNG--VGHINAGIDHVKLINALAKIEPVTLTYH	153
QY	126	DLPOALHDIRGYGLWVNEVQLDFERYARLCFERFGDRVQVMITINXPMWIGAYGATGSN	185
DB	154	DLQALHDIRGLWVNLQIIN-DEFAVAEVCFGKFGGRVKKWITFNEHPTAIGQYDGLQ	212
QY	186	APGRSSINKIST--EGNTATEPWLAKQAQMSHARAVAVSRPFRSQKQIGISLNGDY	243
DB	213	APGRCTILFTLTREBGNSTPEYVGHNVLTATYSDIRKRYKAKQGGSLAIAPDVW	272
QY	244	YEPWDSNEPRDKAERBMEFHGWFPANPIFLKQDYPESKKQLGERLPALTPADPAIIN	303
DB	273	FEP-ESNKTEDIEAQAQDFQGLWFLDPLMF-GDTPSSRSKRVGSLRPLFTSQSSQSLVK	330
QY	304	AGETDPYGNVYTSOPARHLDCGPVETDYLGAIEHQENKDG-----SPVGEBSGL	354

Db 331 -GSLDPVGINHTTYTAARN-----NATNLIGTLHDVADSGVTLLPFKGLSTIGIDRAS 384
QY 355 AMLRSCPMDFRKHLARVYGLYCK-PIYITENGCPGCEENMTCEAVNDPFRIRYFDSHL 413
Db 385 IMLYIVPRGMSRLMNYIKRYGNPPIFITEGMDPNLSLRKALNDAKKIKKHHDYL 444
QY 414 DLSIKRITDDGVVVKYFAMALLDNLNEMSDGCGPRGVTFTDY-TTLKTPPKS 466
Db 445 SSLQASIKEDGCVNKGYPWMSLLDNWEMVAGYSRFGLYFVYDRNLKRYPKDS 498

RESULT 2

T47838
beta-glucosidase-like protein - Arabidopsis thaliana
N/Alternate names: protein T209.120
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
R/Accession: T47838
R/Nyatura, G.; Partam, B.; Dauner, D.; Steer, W.; Holland, R.; Weichselgartner, M.;
submitted to the Protein Sequence Database, February 2000
A/Reference number: Z24475
A/Accession: T47838
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-577 <NTA>
A/Cross-references: EMBL:AL138658
A/Experimental source: Cultivar Columbia, BAC clone T209
C/Genetics:
A/Map position: 3
A/Intons: 45/3; 69/1; 87/3; 113/1; 139/1; 168/2; 293/2; 366/1; 377/3; 412/1; 448/2
A/Note: T209.120
C/Superfamily: Agrobacterium beta-glucosidase

Query Match 37.9%; Score 1003; DB 2; Length 577;
Best Local Similarity 43.7%; Pred. No. 2.5e-71;
Matches 205; Conservative 77; Mismatches 175; Indels 12; Gaps 8;
QY 6 ALPNDFEMGFATAVQIEGAVKEGGSGSIWDYCHLEPSTNGANGVACDHYRDEPD 65
Db 30 SPDDFIFGTASAPQYEGATSEGGKSPITWDHPSLTYPERTKXNDVAVDFYHRYKXD 89
QY 66 FDLTKYGAAYRFSLSRSRIIPGGRLDPVNREGIEFYSKLDLALRGITPMTLYHM 125
Db 90 IKMLKELMDAFRPSISMSRLIPSGKVDGVNKGVEFYKALIDELVANGIEPSMTLYHM 149
QY 126 DLPOALHRYGMLNVEEVDLPERYARLCERFGRDVQVNNITINXPMIOAIYATGSN 185
Db 150 DHPQSLDEYGGFLSPKIVE-DFRDFARICFEFEFGDKVMWTTINEPYIMTVAGYDGNK 208
QY 186 AGRSS--INHGSTEAGTATEPWLAKGAQIMSHARAVAVSRDRPSQKQIGISLNGDY 243
Db 209 AAGRCSSKVNKCCQAGDSSTEPYIVSHHTTLAHAAVAEERKCEKTSHTDQIGIVLSPRW 268
QY 244 YEPWDSNEPRDKEAERMEFHIGMFANPIFLKDYPSMKKQGERLPAITPADFAITLN 303
Db 269 FEPYHSDTDDKEAERLAFELGMLHDPV-IHGDYELIVKVGKGLPSTVEQSCKLQ 327
QY 304 AGEDFYGMNNTYQFARHLDPVPEFDYLGAIHEHQB---NKDGPVEESGLAWLRS 359
Db 328 -NSSDFGGINNYTARFAAHLPIDPEKRFKTDH-IHEWMLKJNSGHIIGGEEGRGLFS 385
QY 360 CPDPRFKHLARVYGLYCK-PIYITENGCPGCEENMTCEAVNDPFRIRYFDSHLDSK 418
Db 386 HPEGLKRVLANIKRYNNMYPYIKENGINDDGTKRREIVKDTFERIEYHKTFFELHK 445
QY 419 AITDGVVVKYFAMALLDNLNEMSDGCGPRGVTFTDY-TTLKTPPKS 466
Db 446 AIVEDGCDVGVYAMSLDNFEMEHGTYARGLYVDFVNGLKRYPKDS 494

RESULT 3

T02401
probable beta-glucosidase At2g44460 [imported] - Arabidopsis thaliana

N/Alternate names: beta-glucosidase homolog F411.27
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C/Accession: T02401; G84878
R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,
submitted to the EMBL Data Library, May 1998
A/Description: Arabidopsis thaliana chromosome II BAC F411 genomic sequence.
A/Reference number: Z14667
A/Accession: T02401

A/Status: translated from GB/EMBL/DDB
A/Molecule type: DNA
A/Residues: 1-577 <ROU>
A/Cross-references: EMBL:AC004521; NID:G3128166; PID:G3128168
A/Experimental source: Cultivar Columbia
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Bentio, M.T.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; Varakken, S.E.; Umayam, L.; Tallon, L.;
euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: G84878
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-577 <STO>
A/Cross-references: GB:AE002093; NID:G3128188; PIDN:AA016092.1; GSPDB:GN00139
C/Genetics:
A/Map position: 2
A/Intons: 48/3; 72/1; 90/3; 116/1; 142/1; 171/2; 256/3; 295/2; 376/1; 409/1; 445/2
C/Superfamily: Agrobacterium beta-glucosidase

Query Match 37.8%; Score 999; DB 2; Length 577;
Best Local Similarity 42.9%; Pred. No. 5.2e-71;
Matches 204; Conservative 78; Mismatches 184; Indels 10; Gaps 8;
QY 8 PNDPFGFRTAAVQIEGAVKEGGSGSIWDYCHLEPSTNGANGVACDHYRDEPD 67
Db 35 PDNFVFGTASAPQYEGATSEGGKSPISWDYSHFPERTRMKNADVAVDFYHRYKXD 94
QY 68 ILTKYGAAYRFSLSRSRIIPGGRLDPVNREGIEFYSKLDLALRGITPMTLYHM 127
Db 95 IKMLKELMDAFRPSISMSRLIPSGKVDGVNKGVEFYKALIDELVANGIEPSMTLYHM 154
QY 128 POALHRYGMLNVEEVDLPERYARLCERFGRDVQVNNITINXPMIOAIYATGSNAP 187
Db 155 PQLSEDEYGGFLSPKIVE-DFRDFSRVCFEEFGDKVMWTTINEPYIMTVAGYDGNKAV 213
QY 188 GRSS--INHGSTEAGTATEPWLAKGAQIMSHARAVAVSRDRPSQKQIGISLNGDY 245
Db 214 GRCSKVNKSCQAGDSSTEPYIVASHHTTLAHAAVAQEF-RKCNKTQDQIGIVLSPLME 272
QY 246 PWDNSNEPRDKEAERMEFHIGMFANPIFLKDYPSMKKQGERLPAITPADFAITLN 305
Db 273 PYDSASPRANENAVKALATELDMHDPV-IHGDYEMKCKLGNLIPSTPEQSCKLK-N 330
QY 306 ETDYFGMNTYQFARHLDPVPEFDYLGAIHEHQBKDSPVGSEGLAWLRS 365
Db 331 SSDFGGINNYTARFAAHLPIDPEKRFKTDH-IHGDYELIVKVGKGLPSTVEQSCKLQ 388
QY 366 KHLARVYGLYCKPI-YITENGCPGCEENMTCEAVNDPFRIRYFDSHLDSK 424
Db 389 KVLNATKDKINPIYIKENGINDDGTKRREIVKDTFERISYHEDHQOQKAIIDBG 448
QY 425 VVVKGYFAMALLDNLNEMSDGCGPRGVTFTDY-TTLKTPPKSALVLMDFAROR 479
Db 449 CVVRGYVWMSLLDNFEMEHGSTRGVVYVDNDNLTPIDDSVWVWFQFLDYKAK 504

RESULT 4

T02402
beta-glucosidase homolog At2g44470 [imported] - Arabidopsis thaliana
N/Alternate names: beta-glucosidase homolog F411.28
C/Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C/Accession: T02402; H84878
R:Rounleay, S.D.; Lin, X.; Kerchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, May 1998
A/Description: Arabidopsis thaliana chromosome II BAC F411 genomic sequence.
A/Reference number: Z14667
A/Accession: T02402
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1591 <R0U>
A/Cross-references: EMBL:AC004521, NID:g3128166; PID:g3128189
A/Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounleay, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
Euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: AB4420; MUID:20083487; PMID:10617197
A/Accession: H84878
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1591 <STO>
A/Cross-references: GB:AE002093; NID:g3128189; PIDN:AA016093.1; GSPDB:GN00139
C/Genetics:
A/Gene: F411.28; At2g44470
A/Map position: 2
A/Intons: 483/; 72/1; 90/3; 116/1; 142/1; 171/2; 256/3; 296/2; 370/3; 416/1; 452/2
C/Superfamily: Agrobacterium beta-glucosidase

Query Match 37.2%; Score 984.5; DB 2; Length 591;
Best Local Similarity 40.0%; Pred. No. 7.5e-70;
Matches 207; Conservative 73; Mismatches 171; Indels 67; Gaps 11;

QY 2 PESLAL-----PNDPEKGFATAVQIEGAVKEGREGSITDTCHELEPRTNGANGVAC 56
DB 24 PESQVDRSSFPDDPVGTAISAFQSEGATSEGGKSTPMDYFSHTPEKTNQMDAVAV 83
QY 57 DHNHRDEDEDLTKGAKAYRFSLSRSRIIPLGGRDLPVNEBIEFYSKLDLALRGI 116
DB 84 DFNHRKXKDDKLIBELNVDAFRFSISNARLIPSGKVDGNGKGVQYKALIDELANGI 143
QY 117 TPWVTLVHMDLPALHNDRYGGMVNEVQDLFERVYALCFERFGDRVQNMITINXPMIOA 176
DB 144 QPSVTLYHMDHPQALDEBYGFLNPQIE-DFRNFAVCFENFGDKYKMMTTINEPVVIS 202
QY 177 IYGYATGNSNAPGSS--INRSTEGNTATPFWLAKQAQINSHARAVAVYRDRPSPQ 224
DB 203 VAGYDYGIRKAVGRCQKVNNSRCQAGDSAIERYIVSHHLLSHAAVQDEFNCKTLDGK 262
QY 235 IGSINGDYEPWDSNEPRDKAEERMEFHIGFANPIFLKKDYPESMKKOLGERLPAI 294
DB 263 IGVVISPMWLEPYDSTSSADKEAVERGLPLELENHLPV-IYGDYPRTEKGVGNRLPAF 321
QY 295 TPADFALNAGETDPFGMNYTSQFARHLDGPVETDYL----- 333
DB 322 TPESQSKL-INSDFPGVNYYSIHFAHL-----PHIDTPRPRPTDHFKEKKCYIYK 376
QY 334 -----CAIHENQENKDSPVGEESGLAMLSCPMFKRLARVYGLYKRP-ITYTEN 384
DB 377 YFSLQDRGKTHSH-----PGLRRLVNIYIDKTNPNPIVYKEN 415
QY 385 GCGPGEENNTCEAVVDPFRIYFDSHLSISKATQDDVAVVYGAFAMALLDLEMSDG 444
DB 416 GIDHYDGTGSRRETIKDTFRISYHODHLKQVHKAILEDGCDVGRVYVMSLFDNFEH 475
QY 445 YGPRFVGTFTDY-TTLKRTPKSKSLVLMFAARQRYK 481
DB 476 YNSRFGYVYDPRKNLQRYPKDSVNWFK-KFLSRPVVR 512

RESULT 5
T09647
beta-glucosidase (EC 3.2.1.21) precursor isoform AH I, cyanogenic - black cherry (fragme

N/Alternate names: amygdalase; celllobiase; gentiobiase
C/Species: Prunus serotina (black cherry)
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C/Accession: T09647
R:Zheng, L.; Poulton, J.E.
Plant Physiol. 109, 31-39, 1995
A/Title: Temporal and spatial expression of amygdalin hydrolase and (R)-(-)-mandelonitril
A/Reference number: Z16795; MUID:96061708; PMID:7480328
A/Accession: T09647
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-551 <CHE>
A/Cross-references: EMBL:U26025; NID:g833834; PIDN:AAA93234.1; PID:g833835
C/Superfamily: Agrobacterium beta-glucosidase
C/Keywords: glycosidase; hydrolase
F:1-23/Domain: signal sequence (fragment) #status predicted <SIG>
F:24-551/Product: beta-glucosidase, cyanogenic #status predicted <MAT>

Query Match 37.2%; Score 983; DB 2; Length 551;
Best Local Similarity 45.3%; Pred. No. 8.9e-70;
Matches 213; Conservative 70; Mismatches 171; Indels 16; Gaps 11;

QY 6 ALPNDEKGFATAVQIEGAVKEGREGSITDTCHELEPRTNGANGVACHTARYDE 64
DB 42 ALPDPGFIFGTASAAVQFEGAAKEDGREGSITDTCHELEPRTNGANGVACHTARYDE 101
QY 65 DFDLTKRYGAKAYRFSLSRSRIIPLGGRDLPVNEBIEFYSKLDLALRGIPTWTLVH 124
DB 102 DVRIKMKMGFDAYRFSISNARLIPSGKVDGNGKGVQYKALIDELANGI 161
QY 125 WDLPOALHNDRYGGMVNEVQDLFERVYALCFERFGDRVQNMITINXPMIOAYGVATGS 184
DB 162 WDLPOALHNDRYGGMVNEVQDLFERVYALCFERFGDRVQNMITINXPMIOAYGVATGS 220
QY 185 NAFGRSSI--NRSTEGNTATPFWLAKQAQINSHARAVAVYRDRPSPQ 242
DB 221 HAPGRCSAMQKLNCTGNSNTEBYLVTHHQLHAAVLYKDEYQASQNGLIGITLVP 280
QY 243 YVEPWSNEPRDKAEERMEFHIGFANPIFLKKDYPESMKKOLGERLPAI 302
DB 281 WEPW-ASEAEEDINAAFRSLDPLFGFPMDEL-TNGVYPLMRSIVGRLNFTBEOGSKL 338
QY 303 NAGETDPFGMNYTSQFARHLDGPVETDYLGAIHENK-----DGSYPGEESGLAMLR 358
DB 339 K-GSPDFIGANTYTTTVAAS--NAPKITSVAASYITTPQVNAATMLKGVPIGMAASGMLY 395
QY 359 SCPDMFRKRLARVYGLYKRP-IYITENGCPCGEENNTCEAVVDPFRIYFDSHLSIS 417
DB 396 VYPRKGIHDLVLYTKEKYNDELITYTENGVDPEFNDPKLSMEALKDTRNIDFYVYRHL 455
QY 418 KAITDQGVVYVYGAFAMALLDLEMSDGGRPVGTFTDY-TLKRTPKS 466
DB 456 AAI-KKGSVKGYFAMVFLDNFEWDAGYTVFGINVYVDNDNKRHSKLS 504

RESULT 6
A57512
beta-glucosidase BG060 precursor - barley
C/Species: Hordeum vulgare (barley)
C/Date: 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change 22-Jun-1999
C/Accession: A57512
R:Leah, R.; Kigel, J.; Svendsen, I.; Mundy, J.
J. Biol. Chem. 270, 15789-15797, 1995
A/Title: Biochemical and molecular characterization of a barley seed beta-glucosidase.
A/Reference number: A57512; MUID:95318171; PMID:7797581
A/Accession: A57512
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-509 <LBA>
A/Cross-references: GB:I41869; NID:g804655; PIDN:AAA87339.1; PID:g804656
C/Genetics: BG060
A/Intons: 100/3; 126/1; 150/1; 179/2; 260/3; 299/2; 372/1
C/Superfamily: Agrobacterium beta-glucosidase

Query Match 36.9%; Score 976; DB 2; Length 509;
Best Local Similarity 43.0%; Pred. No. 2,9e-69;
Matches 206; Conservative 81; Mismatches 170; Indels 22; Gaps 12;

QY 8 PNDFFMGFATAAYOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGDVACDHARYDEDD 67
DB 45 PAFGVGTAAASAYOVEGAMRQGGRCICWDVAALQGMAGNGTADYVDEHARKEDVG 104
QY 68 LUTKVAKAYRPSLSMSRIIPLEGRLDPVNEEGIEFYSKLDALRRGITTPTWTLVHDL 127
DB 105 IMKNMGFDAYRPSISMSRIIPFDG--TGKYNQGVYNNLLIDYMLQQGITPYANNLYHDL 162
QY 128 POALHDYRGMLNVEEVQDPEERARLCPERFGDRYONMTITXPMQAIYATAGSNAP 187
DB 163 PLALHQOYLGLWSPKIVGA-FADYAEFCFKVGDYKWKFTNERVVAALADNGFAP 221
QY 188 GSSSINKSTEGNTATEPMLACKAQIMSHARAVAYSDRFRPSQKQIGISLNGDYEPW 247
DB 222 GRCS--KCPAGGDSSTPEYIVTHNIIILSHAAVQGRREKYOPHQGRIGILLDFVWEP- 278
QY 248 DSENEPRDKEAERREPHIGFANPIFLKQYBESMKQLGERLPLPADPAILNAGET 307
DB 279 HSDTPADDAQAQRAADPHIGFPLDPI-TNGRYPSMLKIVGRPLGFS-ADESERWVKSI 336
QY 308 DFYGNMYTTSOPARHLDG---PVPETD--YLGAIHHEQENKQSPVGEESGLAWLRSCP 361
DB 337 DYVGINQYTSYMKDPGAMNQTPVSGYQDDWYGFYE---RNGVPIOPRANSMDLYIVP 392
QY 362 DMFRGLHARVGLYKRP-IYITENCCPCPGERNMTCEAVNDPFRIRYFSDHLSISKAI 420
DB 393 WGMNAVAYYVERKYNPLMILSENGMDPG--NVSIAQGVHDYAIRYRYDYITELKAI 450
QY 421 TQDGVVVGYPAMALLDNLLEMSDGYGPRFGVTFTDTYTLKRTPKKSALVLDQMFARQ 479
DB 451 DNGARVAGYFAMSLDNLFEWRRLGTARFGIYVVDFTLTKYPKDSALFMKMLSEKR 508

RESULT 7

T02404
probable beta-glucosidase homolog F411.30 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C/Accession: T02404; B84879
R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, May 1998
A/Description: Arabidopsis thaliana chromosome II BAC F411 genomic sequence.
A/Reference number: Z14667
A/Accession: T02404
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-560 <R0U>
A/Cross-references: EMBL:AC004521; NID:g3128166; PIDN:AAIC16095.1; PID:g3128191
A/Experimental source: cultivar Columbia
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: B84879
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-560 <SNO>
A/Cross-references: GB:AE002093; NID:g3128191; PIDN:AAIC16095.1; GSPDB:GN00139
C/Genetics:
A/Gene: At2g44490; F411.30
A/Map position: 2
A/Intons: 3/2; 33/3; 76/3; 102/1; 157/2; 242/3; 281/2; 354/1; 400/1; 438/2
C/Superfamily: Agrobacterium beta-glucosidase

Query Match 36.5%; Score 965; DB 2; Length 560;
Best Local Similarity 42.3%; Pred. No. 2,4e-68;

Matches 204; Conservative 85; Mismatches 169; Indels 24; Gaps 13;

QY 6 ALPNDFFMGFATAAYOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGDVACDHARYDE 64
DB 18 SFPKGLFCTAASSYQIEGAVNCGAQSVMDHPSNFPKIRISDSGNAVVPYHRYKE 77
QY 65 DFDLLTYKAKAYRPSLSMSRIIPLEGRLDPVNEEGIEFYSKLDALRRGITTPTWTLVH 124
DB 78 DIKMKDINNDSFRLSIAMPRLVPGKRDGVSSEGIKFYNDVIDELLANEIRPLVTIFH 137
QY 125 WDLPOALHDYRGMLNVEEVQDPEERARLCPERFGDRYONMTITXPMQAIYATAGS 184
DB 138 WDIPOALDEYGGFLS-EQIIDFRDASLCFERFGDRVSLWCTMNPWYSAGYDGR 196
QY 185 NAFGRSS--INKSTEGNTATEPMLACKAQIMSHARAVAYSDRFRPSQKQIGISLNGD 242
DB 197 KAPORCSKYNVGASVAGMSGYEAYIVSHMMLLHAEEVFF-RKCHINKQIGIAHNL 255
QY 243 YPERWDSNEPRDKEAERREPHIGFANPIFLKQYBESMKQLGERLPLPADPAILN 302
DB 256 WYEPYDSDDDVEGCRNAMDPMGMHQPAC-GDYPETMKKSVGDRLPSTPEQSKL 314
QY 303 NAGETDFYGNMYTTSQFA--RHLDGPVP--ETDYLGAIHHEQENKQSPVGEESGLAWL 357
DB 315 -ISCSYVGINTYSSLFVYSIKGVDPQPTKRTDQ--GYDMKKTINDGQIAKQGSSEWS 371
QY 358 RSCPDMFRKLHARVGLYKRP-IYITENCCPCPGE---ENMTCEAVNDPFRIRYFDS 411
DB 372 FTYPTGIRMLTKYKKTGYNPILLITENGY---GEVAEQSLYMYNPISDTERLEYIEG 428
QY 412 HLDSIKAITQDDVVVVGYPAMALLDNLLEMSDGYGPRFGVTFTDTYTLKRTPKKSALV 470
DB 429 HIAHQALHEQSVREGIYVMSLNDLFEMSGYGVGLYIDYKDGILRRYKMSALWL 488
QY 471 KD 472
DB 489 KE 490

RESULT 8

S78099
furostanol glycoside 26-O-beta-glucosidase F26G - Costus speciosus
C/Species: Costus speciosus
C/Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Jun-2000
C/Accession: S78099; S68626
R/Inoue, K.; Shibuya, M.; Yamamoto, K.; Ebizuka, Y.
submitted to the EMBL Data Library, January 1996
A/Description: Molecular cloning and bacterial expression of a cDNA encoding furostanol
A/Reference number: S78099
A/Accession: S78099
A/Molecule type: mRNA
A/Residues: 1-562 <INO>
A/Cross-references: EMBL:D83177; NID:g1374990; PIDN:BA11831.1; PID:g1374991
A/Experimental source: CSH26d1; rhizome
R/Inoue, K.; Ebizuka, Y.
FEBS Lett. 378, 157-160, 1996
A/Title: Purification and characterization of furostanol glycoside 26-O-beta-glucosidase
A/Reference number: S68626; MUID:96140730; PMID:8549824
A/Accession: S68626
A/Molecule type: protein
A/Residues: 84-86, 'AE', '89-100', 'N', '102-103 <INW>
A/Experimental source: rhizome
A/Complex: heterodimer; 54k chain and 58k chain
C/Function:
A/Description: catalyses the removal of C-26-bound glucose moiety of furostanol glycoside
A/Note: strongly inhibited by diosgenin
C/Superfamily: Agrobacterium beta-glucosidase
C/Keywords: heterodimer

Query Match 36.5%; Score 964.5; DB 2; Length 562;
Best Local Similarity 43.5%; Pred. No. 2,7e-68;
Matches 204; Conservative 78; Mismatches 168; Indels 19; Gaps 12;

QY 6 ALPNDPEMGFATAVOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYHRDE 64
 Db 95 SFPBPGTFGAASAAVOYEGAMNMGGRPSIMDTFTTHDHPKIDHSGDKATOSYKXYKE 154
 QY 65 DFDLLTKYGAAYRFSLSWSRIIPLGRLDPVNEBEGIEFYSKIDALLRRGITPWTLYH 124
 Db 155 DVKLKXLDGDSYRFSISWSRIIPKGTLOGGINOEGIQYVNDLITNELKNGIRPMTLFFH 214
 QY 125 WDLPOLHDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTITNKPMWQAIYGVATGS 184
 Db 215 WDVPQALEDSYKGRSSEIYN-DEKDYADICFEKFGURVHMWTLNPEWLSLTGWVAFGR 273
 QY 185 NAVGRSSINHGSTEGRATEPMLAGKQIMSHAAVAVYSRDFRPSOKGIGISLNDY 244
 Db 274 HAGRCSTWYGCAGADANPEYETHNLLAHANAIVIRDNVYATONGELITNSLWY 333
 QY 245 EPPDSDNEPRKEAERMEHGHIGFANPIFLKDDYPSMKKQJGERLPALTPADFAILNA 304
 Db 334 EPY-SKSHEDVEAATRALDEMFGWMDPL-VNGDYPFIMRALVDRLPFFTHAESELIK- 390
 QY 305 GEDFYGMNYYTSGFAHLDGPV-----PENTDYGAIHEROENKDSPVGEESGLMLRS 359
 Db 391 GSYDFIGINYYTSNYAQH--APVTEHDPONSYFDSYVNOGSEKNGVPIGLQG-SWITYF 447
 QY 360 CPDMFRKHLARVYGLYKGP-IYITENGCPCGSEENMTCEBAVNDPFRIRYFDSHLSISK 418
 Db 448 YPFGKXELLIVKRYCNPKIYITENG---TAEVKEKGYPLHDPKEKVELTHTLAVLQ 504
 QY 419 AITODGVVYGYFAMALLDNLWSDYGPFGVTFYD-TTLKRTPKKS 466
 Db 505 AI-REGVVRVGHFWALTDNFMWDKGYTERFGIYIDYDKNOPKDS 552

RESULT 9

T47836
 beta-glucosidase-like protein - Arabidopsis thaliana
 N/Alternate names: protein T209.100
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
 C/Accession: T47836
 R/Author: G.; Fartmann, B.; Dauner, D.; Steer, W.; Holland, R.; Weichselgartner, M.;
 submitted to the Protein Sequence Database, February 2000
 A/Reference number: 224475
 A/Accession: T47836
 C/Genetic: preliminary
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residue: 1-534 <N>A>
 A/Cross-references: EMBL:ALJ38658
 A/Experimental source: cultivar Columbia; BAC clone T209
 A/Map position: 3
 A/Intons: 51/1; 95/1; 121/1; 150/2; 235/3; 274/2; 347/1; 393/1; 429/2
 A/Note: T209.100
 C/Superfamily: Agrobacterium beta-glucosidase

Query Match 36.3%; Score 960.5; DB 2; Length 534;
 Best Local Similarity 42.8%; Pred. No. 5.2e-66;
 Matches 207; Conservative 75; Mismatches 185; Indels 17; Gaps 11;
 QY 8 PNDPEMGFATAVOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRDE 67
 Db 14 PEGFLFTASSAVOYEGABNAPRGESVMDTFVAKYPERNCYSNADQAIIFYNHYKDDIQ 73
 QY 68 LITRYGAKAYRFSLSWSRIIPLGRLDPVNEBEGIEFYSKIDALLRRGITPWTLYH 127
 Db 74 RMDKINNDARFISMSRIPPLGKSKSVNKEGQIFVNDLIDELANGITPLATLFFMDT 133
 QY 128 POLHDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTITNKPMWQAIYGVATGS 187
 Db 134 POLHDEYSGFLSEAVD-DEKDPALCFEFGGRVAKWTLNPEWLSLTGWVAFGR 192
 QY 188 GRSS--INKHSTEGNTATEPMLAGKQIMSHAAVAVYSRDFRPSOKGIGISLNDY 245

Db 193 GRASKYMNBAVAVGESGLEEYTVSHNLLAHAAVEVF-RNNCKDKGIGIANCPWFE 251
 QY 246 PMSNPEPRDEAERMEHGHIGFANPIFLKDDYPSMKKQJGERLPALTPADFAILNAG 305
 Db 252 PYDSNCPKDI EAERMEHGHIGFANPIFLKDDYPSMKKQJGERLPALTPADFAILNAG 309
 QY 306 ETDYFGMNYTSGFAHLDGPV-----DGPVETDYGAIHEROENKDSPVGEESGLMLRS 360
 Db 310 SFPBPGTFGAASAAVOYEGAMNMGGRPSIMDTFTTHDHPKIDHSGDKATOSYKXYKE 367
 QY 361 PDMFRKHLARVYGLYKGP-IYITENGCPCGSEENMTCEBAVNDPFRIRYFDSHLSISK 418
 Db 368 PGLRKFVLYAKKYS PKEMITENGCPCGSEENMTCEBAVNDPFRIRYFDSHLSISK 426
 QY 419 AITODGVVYGYFAMALLDNLWSDYGPFGVTFYD-TTLKRTPKKSALVLCDFE 477
 Db 427 AITODGVVYGYFAMALLDNLWSDYGPFGVTFYD-TTLKRTPKKSALVLCDFE 486
 QY 478 QRYK 481
 Db 487 EEIE 490

RESULT 10

T02403
 probable beta-glucosidase At2g44480 [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 02-Feb-2001
 C/Accession: T02403; A84879
 R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
 Wuse, D.; Nierman, W.C.; White, O.; Bisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A/Reference number: A84420; NMID:20083487; PMID:10617197
 A/Accession: T02403
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residue: 1-517 <ROU>
 A/Cross-references: EMBL:AC004521; NID:G3128166; PIDN:AAIC6094.1; PID:G3128190
 A/Experimental source: cultivar Columbia
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
 Wuse, D.; Nierman, W.C.; White, O.; Bisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A/Reference number: A84420; NMID:20083487; PMID:10617197
 A/Accession: A84879
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residue: 1-517 <STO>
 A/Cross-references: GB:AE002093; NID:G3128190; PIDN:AAIC6094.1; GSPDB:GN00139
 C/Genetic: preliminary
 A/Map position: 2
 A/Intons: 54/3; 78/1; 97/3; 123/1; 149/1; 178/2; 262/3; 301/2; 374/1; 384/3; 419/1; 455/1
 C/Superfamily: Agrobacterium beta-glucosidase

Query Match 35.7%; Score 944.5; DB 2; Length 517;
 Best Local Similarity 41.1%; Pred. No. 9.1e-67;
 Matches 197; Conservative 91; Mismatches 160; Indels 31; Gaps 11;
 QY 6 ALPNDPEMGFATAVOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYHRDE 64
 Db 39 SFPDPRFGAASAAVOYEGABNAPRGESVMDTFVAKYPERNCYSNADQAIIFYNHYKDDIQ 98
 QY 65 DFDLLTKYGAAYRFSLSWSRIIPLGRLDPVNEBEGIEFYSKIDALLRRGITPWTLYH 124
 Db 99 DVAMKEIGDSEFRSISWSRIIPRGTVAGVQAQINPNNHINELISNGIRPMTLFFH 158
 QY 125 WDLPOLHDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTITNKPMWQAIYGVATGS 184
 Db 159 WDVPQALEDSYKGRSSEIYN-DEKDYADICFEKFGURVHMWTLNPEWLSLTGWVAFGR 217
 QY 185 NAPGR--SIINKHSTEGNTATEPMLAGKQIMSHAAVAVYSRDFRPSOKGIGISLNDY 243

```
Db      218 IAPGRCSSVYVQCTVGNATPEPYLVAHLYLISHAATVQYRREKYSFHGTTGMITQYTW 277
      244 YEPNMSNPRDKEAERKMEFHIGFANPFIPLKKQYPSKMKQCEBRLPALTPADPAILN 303
      Db      278 MTP-KYNTFPACREAAKRRALDFFFGWADPFI-TVGYGPKTMRBLVGNRLPKFTKXSKXVR 335
      Qy      304 AGETDPYGMNYYTSQFARHL-----DGVPEPTDYLGAHHEQNDKSPVGE 351
      Db      336 -GSFPFGLNYYTSRYVEDVMFYATNTNLSYTTDSRYNOT-----TEKNQVPSBP 384
      Qy      332 SGLAMLRSCPDMEFRKLARVGLYKRP-IYTTENGCPGCEENMTCEAVNDPFRIRYD 410
      Db      385 TSADWLFTCPCEGFQDVLVIYKSKFQNPVILYTENGMPSENDKSLSNVIALNDEAKIKYHQ 444
      Qy      411 SHLDSISYATDQGVVGVFAMALLDNLNEMSDGSPRGVTFMDYTT-LKRTPKSKL 468
      Db      445 LHLTLLEAVSO-GADVKGYYTWSLMDPEWMEFGYKRYGLVYDFQDGLKRHLKSSAL 502
```

RESULT 11

```
T09657
beta-glucosidase (EC 3.2.1.21) precursor, cyanogenic - black cherry
N:Alternate names: amygdalase; cellobiase; gentiobiase
C:Species: Prunus serotina (black cherry)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 26-Aug-1999
Accession: T09657
R:Poulton, J.E.; Jurrk, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z16802
A:Accession: T09657
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-549 <POU>
A:Cross-references: EMBL:U50201; NID:G1236960; PID:G1236961
C:Superfamily: Agrobacterium beta-glucosidase
C:Keywords: glycosidase; hydrolase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-549/Product: beta-glucosidase, cyanogenic #status predicted <MAT>
```

```
Query Match      35.7%; Score 943.5; DB 2; Length 549;
Best Local Similarity 42.5%; Pred. No. 1.2e-66;
Matches 198; Conservative 84; Mismatches 169; Indels 15; Gaps 11;

Qy      11 FEMGPAATAVQIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYRDEFDLL 69
      Db      54 FTFGAATAAYQLEGANIDGRPSVDMFTHEHEPKITDGSNGVALDOYHRYKEDVALM 113
      Qy      70 TKYGAKAARSLMSRIIPLGRLDPVNEBGEIEFSKIDALLRGITTPWVTLYHMDLPQ 129
      Db      114 KDMGLDAVRFISWSRLLPNTLSGGINKGIEYYNNLTNELLNNGIEPLVTLFHHMDVPQ 173
      Qy      130 ALHRYGGLMVAEEVQDPERFARLCFERFGDRVONMTTINKPMIOIYGAATGSNAQR 189
      Db      174 ALVVEYDGLSPRIYD-DFAVYANLCYERFDRYKAMTTLNEPFTVSNHGTTGITHAPGR 232
      Qy      190 SSINKHST--EGNTATEPMLAGKAQIMSHARAVAVYSDRFPSSQKQIGISLNGDYEPW 247
      Db      233 CSCWDPCTLCGDSGTBEYLVTHHLLAHAAAVKLTYREKYQASONGVIGITIVSHWEP- 291
      Qy      248 DSNPRDKEAERKMEFHIGFANPFIPLKKQYPSKMKQCEBRLPALTPADPAILNAGET 307
      Db      292 ASBESQDDKDAARLADDFYGMFMEPL--TRGDYPTQMSIYGSRLPNFTBESQSKLN-GSY 349
      Qy      308 DFGYNNYYTSQFA-----RHLDGPVPEYDYLGA-IHEQNDKSPVGEESGLAMLRSCPD 362
      Db      350 DYIGVNYYSARVAYATNNVSVFPPPSVATDAYVNTTDLNGVPIRQASDMLYYTPK 409
      Qy      363 MFRGHARVGLYKRP-IYTTENGCPGCEENMTCEAVNDPFRIRYFDSHLSISAIT 421
      Db      410 GLYDLVLYTEKRYNDPMVYITENGMDEFNNPKLSLEBALDANRDIYYRHLCYLQAAI- 468
      Qy      422 QDGVVGVGVFAMALLDNLNEMSDGSPRGVTFMDYTT-LKRTPKKS 466
```

```
Db      469 KEGANVOGYFAMSLDNLNEMSEGYTVFAGINYIDYDNGLERHSKLS 514
```

RESULT 12

```
S50756
beta-D-glucosidase - oat
C:Species: Avena sativa (oat)
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 22-Jun-1999
Accession: S50756
R:Gus-Mayer, S.; Brunner, H.; Schneider-Poetsch, H.A.W.; Ruediger, W.
plant Mol. Biol. 26, 909-921, 1994
A:Title: Avenacosidase from oat: purification, sequence analysis and biochemical character
A:Accession: S50756; MUID:95093031; PMID:8000004
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-574 <GUS>
A:Cross-references: EMBL:X78433; NID:G505278; PID:CAAS196.1; PID:G505279
C:Superfamily: Agrobacterium beta-glucosidase
```

```
Query Match      35.2%; Score 932; DB 2; Length 574;
Best Local Similarity 40.9%; Pred. No. 1e-65;
Matches 202; Conservative 86; Mismatches 182; Indels 22; Gaps 14;
```

```
Qy      1 MPESIALPNDFEMGPAATAVQIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHY 59
      Db      68 VPKRDMFPPEFMGAASAAVQIEGANNEGKSSMDNFCHSPIDKIMDSNDVAAASTY 127
      Qy      60 HRYDEDFDLLTKYGAKAARPSLSWSRIIPLGRLDPVNEBGEIEFSKIDALLRGITPW 119
      Db      128 YMYKEDVRMLKEIGMDSYRFSISMPRLPKGTLDDGINHEGIQYNDLDCLENGIKFY 187
      Qy      120 VTLXHMDDLPAALDRYGMNVAEVQDPERFARLCFERFGDRVONMTTINKPMIOIYGA 179
      Db      188 ITLFHMDTPQALADEKDFLD-RIIVKDYDYATVCFEHHGDKYKMWFTFNEHPSFCGLG 246
      Qy      180 YATGSNAPG-RSSINKHST--EGNTATEPMLAGKAQIMSHARAVAVYSDRFPSSQKQIG 236
      Db      247 YGTGHAHAPGARCSAGMTCVIPEEDALRNPIYVHNLLAHAEVVDYNNK-FYSGDGOIG 305
      Qy      237 ISLNGDYEPWSDNEPRDKEAERKMEFHIGFANPFIPLKKQYPSKMKQCEBRLPALTP 296
      Db      306 MVLDVNAVEPYGNPF-LDQOQERAI DFHIGWLEBPM-VRGDVPFSMRSLVGRRLPFTFK 363
      Qy      297 ADAFALLNAGETDPYGMNYYTSQFARLD-----GVPPEPTDYLGAHHEQND- 349
      Db      364 SEQPKL-VSSYDVGINYTSRPAKIDISPEFPIKINTD--DVSNEPVDNSNGIPIG 419
      Qy      350 EESGLAMLRSCPDMEFRKLARVGLYKRP-IYTTENG-CPGCEENMTCEAVNDPFRIR 407
      Db      420 PDVGMFYIYSYRGLNILLRMKEKGNPIYITTEGTADMOMGMPMTDPLDDPLRIE 479
      Qy      408 YFPHSLDSISKAITQDGVVGVFAMALLDNLNEMSDGSPRGVTFMDYTT-LKRTPKKS 466
      Db      480 YLQOHTMTAIKEADLGRRLRGHFTSLIDNEMSGYLSRFGIYVIDRNDGCKRIMKS 539
      Qy      467 ALVIXDMFAPARORV 480
      Db      540 AKWLKEPFGATKKL 553
```

RESULT 13

```
JC7539
beta-glucosidase (EC 3.2.1.21) - Thai rosewood
N:Alternate names: dalcocchinase; dalcocchinin-8'-O-beta-glucoside beta-glucosidase
C:Species: Dalbergia cochinchinensis
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
Accession: JC7539; PC7111
R:Calims, J.R.K.; Champatenechai, V.; Sriomsap, C.; Wiltman-Liebold, B.; Thiede, B.; St
J. Biochem. 128, 999-1008, 2000
A:Title: Sequence and expression of Thai rosewood beta-glucosidase/beta-fucosidase, a far
A:Reference number: JC7539; MUID: 21062552; PMID:11098143
```

A:Contents: Seeds
A:Accession: Jc7539
A:Molecule type: mRNA
A:Residues: 1-547 <CAI>
A:Cross-references: GB:AF163097
A:Accession: PC7111
A:Molecule type: protein
A:Residues: 66-74;81-95;100-104;106-112;127-141;138-351;410-422;465-472;500-506;507-516
C:Comment: This enzyme belongs to family 1 glycosyl hydrolase glycoproteins and plays a role in degradation, and defense against herbivores. It has aglycone specificity for isoflavonone
C:Superfamily: Agrobacterium beta-glucosidase
C:Keywords: glycoprotein; glycosidase; growth regulation; hydrolase

Query Match	35.1%	Score 928;	DB 2;	Length 547;
Best Local Similarity	42.1%	Pred. NO. 2e-65;		
Matches 204;	Conservative	78;	Mismatches 173;	Indels 30;
			Gaps	12;

```

QY      8 PNDEMEFPATAYOIEGAVKEGGRPSIMDYCHLEPSR--TNGANGVACIHYHRXEDF 66
Db      46 PSDDIFGFIASSYQYEGE-----GRVPSIMDFTHQYEPKIMDRBNGDVANDQPHRYKKDI 101
QY      67 DLTITGAKAVRPSLSRSRIIPUGGRLDPVNVEGIEFYKSIDALLRRGITPMVTLVYMD 126
Db      102 AIMDMNLDAYRMSISWPRIIPTRGVSGGIQGTGDVYNNRLINESLANGITPPTIIFMD 161
QY      127 LPQALHDRVGMVLNVEEYODFEERYARLCFEPFGDRVQNMVTTIXPMIQAIVYATGSNA 186
Db      162 LPQALHEDEYGFNL-HGVANDFQDYADLCFQDFGRVGMVTTINLPSIFTMANGVAYGMA 220
QY      187 PGR--SSINKHSTEGNATEBPWLAKQAIMSHADAVALVYSHDRPPSQKGQIGISLNGDY 244
Db      221 PGRCSPPSNPCTCGDAGTETVYLVANHLLISHAATVOYKKKYKYEHOXGTTIGISLHVWV 280
QY      245 EPMCSNEPRDKEAERMEFHIQWFPANDIPFLKKOYPSMKKKQGLEPRLPALRPADPATINA 304
Db      281 IPL-SNSTSDONAKORYIYDFTCGMFMDPLTAGR-YPDSMQVLVGDRLPKFT-TQAKLYK 337
QY      305 GETDFYGMNNTYQSOFARHLDGPV-----PETDYLGAIHEQNKDGS PVGEESGLA 355
Db      338 GSPFPIGLINTYTTNVTATKSDASTCCPSYLTDPQVTL-----QQRNQGVIFIPVPSG 390
QY      356 WLRSCPMFRGHILARVYGLYKRP-IYITENCCPCRGEEHMTCEEAVNDPRIRAFDHLID 414
Db      391 WMCITPGRGLRLLYPFEKXKNNPLVYITENGI DEKNDKSSLSLESLLDYTAIDSYRHLF 450
QY      415 SISKATIQDGVVVKGFAMALLDNLWESDGYGPRPFGVTFDYTTILKTRPKKSALVLXDMF 474
Db      451 YVRVAI-RSGANVXGFFAMSLDDNFEMWEGYTSRFGLYFVYNTTLNRPKLSATMFK-YF 508
QY      475 AAROR 479
Db      509 LARDO 513

```

```

RESULT 14
S45723
P60 protein - oat
C:Species: Avena sativa (oat)
C:Date: 10-Dec-1994 #sequence_revision 19-Apr-1996 #text_change 17-Mar-1999
C:Accession: S45723
R:Gus-Meyer, S.; Brunner, H.; Schneider-Poetsch, H.A.W.; Lotespelch, F.; Eckerskorn, C.;
FEBS Lett. 347, 51-54, 1994
A:Title: The amino acid sequence previously attributed to a protein kinase or a TCPI-rel
A:Reference number: S45723; MUID:94283628; PMID:8013661
A:Accession: S45723
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-519 <GUS>
A:Superfamily: Agrobacterium beta-glucosidase

```

Query Match	35.1%	Score 927	DB 2	Length 519
Best Local Similarity	40.7%	Pred. No. 2.2e-65		
Matches 201	Conservative 89	Mismatches 182	Indels 22	Gaps 14

[illegible]

RESULT 15

Probable beta-glucosidase [imported] - Arabidopsis thaliana
 C.Species: Arabidopsis thaliana (mouse-ear cress)
 C.Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C.Accession: T02400; F84878
 C.Round(s): S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, R. submitted to the EMBL data library, May 1998
 A.Description: Arabidopsis thaliana chromosome II BAC F411 genomic sequence.
 A.Reference number: Z14667
 A.Accession: T02400
 A.Status: translated from GB/EMBL/DBDJ
 A.Molecule type: DNA
 A.Residues: 1-506 <ROU>
 A.Cross-references: EMBL:AC004521; NID:g3128166; PIDN:AACT16091.1; PID:g3128187
 A.Experimental source: cultivar Columbia
 R.Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Nuss, D.; Moeffar, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; Ems, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-766, 1999
 A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A.Reference number: A84420; MUID:20083487; PMID:10617197
 A.Accession: F84878
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-506 <STO>
 A.Cross-references: GB:AEO02093; NID:g3128187; PIDN:AACT16091.1; GSPDB:GN00139
 C.Genetics:
 A:Gene: F411.26; AC2g44450
 A:Map position: 2
 A:Insertions: 50/3; 74/1; 93/3; 119/1; 145/1; 174/2; 259/3; 370/1; 380/3; 415/1; 447/2
 C:Superfamily: Agrobacterium beta-glucosidase

Query Match	35.0%	Score 925.5;	DB 2;	Length 506;
Best Local Similarity	43.1%	Pred. No. 2.8e-65;		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2004, 15:45:11 ; Search time 60 Seconds
(without alignments)
2279.217 Million cell updates/sec

Title: US-10-026-140-2

Perfect score: 2644
Sequence: 1 MPESLALPNDPFWGFATYAA.....KSLVLKDNFAARQRYKVA 484

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2629	99.4	484	7	ADD22903 T. reesei
2	1037	39.2	514	3	AA828767 Arabidops
3	1037	39.2	516	3	AA828766 Arabidops
4	1037	39.2	523	3	AA828765 Arabidops
5	981.5	37.1	507	3	AA826813 Beta-Prim
6	965	36.5	549	3	AA829189 Arabidops
7	965	36.5	560	3	AA829188 Arabidops
8	965	36.5	560	3	ADCO3121 A. thalia
9	943.5	35.7	545	6	ABP71306 Prunus se
10	939	35.5	512	3	AA814175 Arabidops
11	939	35.5	521	3	AA814174 Arabidops
12	930	35.2	495	5	ABP99404 Arabidops
13	925.5	35.0	506	3	AA829349 Arabidops
14	923.5	34.9	657	3	AA852193 Beta gluc
15	923	34.9	507	3	AA850272 Arabidops
16	920	34.8	487	3	AA849752 Arabidops
17	910	34.4	513	4	AA842119 Coniferin
18	905	34.2	574	4	AB876821 Avena sat
19	880.5	33.3	459	5	AA826442 White clo
20	874.5	33.1	555	3	AA813434 C. roseus
21	865.5	32.7	524	3	AA826437 Arabidops
22	865	32.7	567	4	AA812298 Human PRO
23	865	32.7	567	5	AB884990 Human PRO
24	865	32.7	567	5	AB832377 Novel hum
25	865	32.7	567	5	AB895596 Human ang

26	865	32.7	567	6	AB017742	Ab017742	Novel hum
27	865	32.7	567	6	ABU80996	Abu80996	Human PRO
28	865	32.7	567	6	ABU66696	Abu66696	Human PRO
29	865	32.7	567	6	ABU59777	Abu59777	Novel sec
30	865	32.7	567	6	ABO24967	Ab024967	Human sec
31	865	32.7	567	6	ABU66972	Abu66972	Human sec
32	865	32.7	567	6	ADA45773	Ada45773	Novel hum
33	865	32.7	567	6	ADA76204	Ada76204	Human PRO
34	865	32.7	567	6	ADA18854	Ada18854	Human PRO
35	865	32.7	567	6	ADA61477	Ada61477	Human PRO
36	865	32.7	567	6	ADA619262	Ada619262	Novel hum
37	865	32.7	567	6	ADA827803	Ada827803	Human PRO
38	865	32.7	567	6	ADA86282	Ada86282	Novel hum
39	865	32.7	567	6	ADA815846	Ada815846	Human PRO
40	865	32.7	567	6	ADA47632	Ada47632	Human PRO
41	865	32.7	567	6	ADA67427	Ada67427	Human PRO
42	865	32.7	567	6	ADA30434	Ada30434	Human PRO
43	865	32.7	567	6	ADA85730	Ada85730	Novel hum
44	865	32.7	567	6	ADA96942	Ada96942	Human PRO
45	865	32.7	567	6	ADA79246	Ada79246	Human PRO

ALIGNMENTS

RESULT 1
ID ADD22903 standard; protein; 484 AA.

AC ADD22903;

DT 15-JAN-2004 (first entry)

DE T. reesei BGL5 beta-glucosidase protein.

XX BGL5; beta-glucosidase; fungus; ethanol; sugar;
XX fermentative microorganism; endoglucanase; cellbiohydrolyase; wine making;
KW aroma; detergent; softening agent; cotton; wood pulp; enzyme.

XX OS Hypocrea jecorina.

FH Key Location/Qualifiers

FT MISC-difference 13 /note= "Encoded by TGC"

FT MISC-difference 164 /note= "Encoded by CAC"

FT MISC-difference 171 /note= "Encoded by GAW"

FT MISC-difference 389 /note= "Encoded by CT"

FT MISC-difference 407 /note= "Encoded by GGS"

PN US2003114330-A1.

PD 19-JUN-2003.

PF 18-DEC-2001; 2001US-00026140.

PR 18-DEC-2001; 2001US-00026140.

PA (DUNN/) DUNN-COLEMAN N.

PA (GOED/) GOEDEGEBUUR F.

PA (WARD/) WARD M.

PA (YAOJ/) YAO J.

XX Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;

XX WPI; 2003-810909/76.

XX DR N-PSDB; ADD22902, ADD22904.

XX PT New substantially purified BGL5 polypeptide with the biological activity

XX of beta-glucosidase, useful for degrading biomasses to ethanol.

XX Claim 18; Fig 2; 21pp; English.

XX The invention discloses a substantially purified BGL5 polypeptide with
 XX the biological activity of a beta-glucosidase. Also claimed is an
 XX isolated polynucleotide derived from a fungal source, which comprises a
 XX nucleotide sequence encoding an enzyme having beta-glucosidase activity,
 XX a Trichoderma reesei bgl5 nucleic acid coding sequence, or its
 XX complement, an expression construct including the T.reesei bgl5 cDNA
 XX sequence, a purified recombinant enzyme having beta-glucosidase activity,
 XX a recombinant host cell comprising a deletion or insertion or other
 XX alteration in the bgl5 gene which inactivates the gene and prevents BGL5
 XX polypeptide production, an antisense oligonucleotide, a detergent
 XX composition comprising the polypeptide and expressing a heterologous
 XX polypeptide having beta-glucosidase activity in an Aspergillus sp. BGL5
 XX is useful for producing ethanol, which involves contacting a biomass
 XX composition with an enzymatic composition comprising the polypeptide to
 XX yield a sugar solution, adding to the sugar solution a fermentative
 XX microorganism and culturing the fermentative microorganism under
 XX conditions sufficient to produce ethanol, where the biomass composition
 XX may be optionally pretreated. The method further involves the addition of
 XX at least one endoglucanase or cellobiohydrolase. The pretreatment is with
 XX a dilute acid. BGL5 is useful in wine making for enhancing the potential
 XX aroma of the finished wine product. The detergent is useful as softening
 XX agent and also useful for improving the feel of cotton fabrics and for
 XX degrading wood pulp into sugars. The sequence presented is the T. reesei
 XX BGL5 protein.

XX Sequence 484 AA;

Query Match 99.4%; Score 2629; DB 7; Length 484;
 Best Local Similarity 99.8%; Pred. No. 7e-254;
 Matches 483; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPESALPNDEMGFATAAYOIEGAVKEGSGPSINPTYCHLEPSRTKANGADYACDHYH 60
 DB 1 MPESALPNDEMGFATAAYOIEGAVKEGSGPSINPTYCHLEPSRTKANGADYACDHYH 60
 QY 61 RYDEDFDLITKYGAKAVRFSLSWSRIIPDGRDLPVNEBEGIEFYSKIDALLRGITPMW 120
 DB 61 RYDEDFDLITKYGAKAVRFSLSWSRIIPDGRDLPVNEBEGIEFYSKIDALLRGITPMW 120
 QY 121 TLVHMDLPQALHNDRYGGMWLANVEYQVLDPERYARLCFERFGDGVONWITINXPMIOAIYGY 180
 DB 121 TLVHMDLPQALHNDRYGGMWLANVEYQVLDPERYARLCFERFGDGVONWITINXPMIOAIYGY 180
 QY 181 ATGSNAPGRSSINHGSTEAGTATEPMLAGKAQIMSHARAVVSRDPRPSQKOTIGISLN 240
 DB 181 ATGSNAPGRSSINHGSTEAGTATEPMLAGKAQIMSHARAVVSRDPRPSQKOTIGISLN 240
 QY 241 GDYIEPMDSNRPDKAERMEFHIGWFPANPIFLKDYEPESMKQUGERLPALTTPADFA 300
 DB 241 GDYIEPMDSNRPDKAERMEFHIGWFPANPIFLKDYEPESMKQUGERLPALTTPADFA 300
 QY 301 ILNNGETDFYGMNYYTQSFARHLDGPVPEVDYLCAIHEQGNKDGSPVGESEGLAMTRSC 360
 DB 301 ILNNGETDFYGMNYYTQSFARHLDGPVPEVDYLCAIHEQGNKDGSPVGESEGLAMTRSC 360
 QY 361 PDMERKHLARVYGLYKPRITYTENGCCPCPEENMTCEAVNDPRIRIYFPSSHLDSTSKAI 420
 DB 361 PDMERKHLARVYGLYKPRITYTENGCCPCPEENMTCEAVNDPRIRIYFPSSHLDSTSKAI 420
 QY 421 TODGVVVKGFAYMALLDNLEWSDGYGRFGVTFDTYTLTKRTPKKSLVLKMPFAARQRY 480
 DB 421 TODGVVVKGFAYMALLDNLEWSDGYGRFGVTFDTYTLTKRTPKKSLVLKMPFAARQRY 480
 QY 481 KVAA 484
 DB 481 KVAA 484

RESULT 2
 AAG28767

ID AAG28767 standard; protein; 514 AA.
 XX AAG28767;
 AC 17-OCT-2000 (first entry)
 DT Arbidopsis thaliana protein fragment SEQ ID NO: 34109.
 DE Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 OS Arbidopsis thaliana.
 EN EPI033405-A2.
 PD 06-SEP-2000.
 PF 25-FEB-2000; 2000EP-00301439.
 XX 25-FEB-1999; 99US-0121825P.
 XX 05-MAR-1999; 99US-0123180P.
 XX 09-MAR-1999; 99US-0123548P.
 XX 23-MAR-1999; 99US-0125788P.
 XX 25-MAR-1999; 99US-0126264P.
 XX 29-MAR-1999; 99US-0126785P.
 XX 01-APR-1999; 99US-0127462P.
 XX 06-APR-1999; 99US-0128234P.
 XX 08-APR-1999; 99US-0128714P.
 XX 16-APR-1999; 99US-0129845P.
 XX 19-APR-1999; 99US-0130077P.
 XX 21-APR-1999; 99US-0130449P.
 XX 23-APR-1999; 99US-0130510P.
 XX 28-APR-1999; 99US-0130891P.
 XX 30-APR-1999; 99US-0131449P.
 XX 30-APR-1999; 99US-0132048P.
 XX 04-MAY-1999; 99US-0132407P.
 XX 05-MAY-1999; 99US-0132484P.
 XX 06-MAY-1999; 99US-0132485P.
 XX 06-MAY-1999; 99US-0132486P.
 XX 07-MAY-1999; 99US-0132487P.
 XX 11-MAY-1999; 99US-0132863P.
 XX 14-MAY-1999; 99US-0134256P.
 XX 14-MAY-1999; 99US-0134218P.
 XX 14-MAY-1999; 99US-0134219P.
 XX 14-MAY-1999; 99US-0134221P.
 XX 14-MAY-1999; 99US-0134370P.
 XX 18-MAY-1999; 99US-0134768P.
 XX 19-MAY-1999; 99US-0134941P.
 XX 20-MAY-1999; 99US-0135124P.
 XX 21-MAY-1999; 99US-0135353P.
 XX 25-MAY-1999; 99US-0135629P.
 XX 27-MAY-1999; 99US-0136021P.
 XX 28-MAY-1999; 99US-0136392P.
 XX 01-JUN-1999; 99US-0136782P.
 XX 03-JUN-1999; 99US-0137222P.
 XX 04-JUN-1999; 99US-0137528P.
 XX 07-JUN-1999; 99US-0137502P.
 XX 08-JUN-1999; 99US-0137724P.
 XX 10-JUN-1999; 99US-0138094P.
 XX 10-JUN-1999; 99US-0138540P.
 XX 14-JUN-1999; 99US-0138847P.
 XX 16-JUN-1999; 99US-0139119P.
 XX 17-JUN-1999; 99US-0139452P.
 XX 17-JUN-1999; 99US-0139453P.
 XX 18-JUN-1999; 99US-0139454P.
 XX 18-JUN-1999; 99US-0139455P.
 XX 18-JUN-1999; 99US-0139456P.
 XX 18-JUN-1999; 99US-0139457P.
 XX 18-JUN-1999; 99US-0139458P.
 XX 18-JUN-1999; 99US-0139459P.
 XX 18-JUN-1999; 99US-0139460P.

PR	18-JUN-1999	99US-0139461P
PR	18-JUN-1999	99US-0139462P
PR	18-JUN-1999	99US-0139463P
PR	18-JUN-1999	99US-0139750P
PR	18-JUN-1999	99US-0139763P
PR	21-JUN-1999	99US-0139817P
PR	22-JUN-1999	99US-0139891P
PR	23-JUN-1999	99US-0140353P
PR	24-JUN-1999	99US-0140354P
PR	24-JUN-1999	99US-0140655P
PR	28-JUN-1999	99US-0140823P
PR	29-JUN-1999	99US-0140991P
PR	30-JUN-1999	99US-0141287P
PR	01-JUL-1999	99US-0141842P
PR	01-JUL-1999	99US-0142154P
PR	02-JUL-1999	99US-0142035P
PR	02-JUL-1999	99US-0142350P
PR	08-JUL-1999	99US-0142803P
PR	09-JUL-1999	99US-0142920P
PR	12-JUL-1999	99US-0142977P
PR	13-JUL-1999	99US-0143542P
PR	14-JUL-1999	99US-0143624P
PR	15-JUL-1999	99US-0144005P
PR	16-JUL-1999	99US-0144085P
PR	16-JUL-1999	99US-0144086P
PR	19-JUL-1999	99US-0144325P
PR	19-JUL-1999	99US-0144332P
PR	19-JUL-1999	99US-0144333P
PR	19-JUL-1999	99US-0144334P
PR	19-JUL-1999	99US-0144335P
PR	20-JUL-1999	99US-0144352P
PR	20-JUL-1999	99US-0144632P
PR	20-JUL-1999	99US-0144684P
PR	21-JUL-1999	99US-0144841P
PR	21-JUL-1999	99US-0145086P
PR	21-JUL-1999	99US-0145088P
PR	22-JUL-1999	99US-0145085P
PR	22-JUL-1999	99US-0145087P
PR	22-JUL-1999	99US-0145102P
PR	22-JUL-1999	99US-0145128P
PR	23-JUL-1999	99US-0145145P
PR	23-JUL-1999	99US-0145218P
PR	23-JUL-1999	99US-0145224P
PR	26-JUL-1999	99US-0145276P
PR	27-JUL-1999	99US-0145591P
PR	27-JUL-1999	99US-0145591P
PR	28-JUL-1999	99US-0145591P
PR	02-AUG-1999	99US-0146386P
PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146389P
PR	02-AUG-1999	99US-0146389P
PR	04-AUG-1999	99US-0147204P
PR	04-AUG-1999	99US-0147302P
PR	05-AUG-1999	99US-0147193P
PR	05-AUG-1999	99US-0147260P
PR	06-AUG-1999	99US-0147303P
PR	06-AUG-1999	99US-0147451P
PR	09-AUG-1999	99US-0147453P
PR	09-AUG-1999	99US-0147955P
PR	10-AUG-1999	99US-0148117P
PR	11-AUG-1999	99US-0148319P
PR	12-AUG-1999	99US-0148341P
PR	13-AUG-1999	99US-0148565P
PR	13-AUG-1999	99US-0148684P
PR	16-AUG-1999	99US-0149358P
PR	17-AUG-1999	99US-0149317P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149723P
PR	20-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149920P

PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156468P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match 39.2%; Score 1037; DB 3: Length 514;
 Best Local Similarity 44.1%; Pred. No. 2e-94;
 Matches 209; Conservative 85; Mismatches 156; Indels 24; Gaps 10;

```

OY      6 ALPDPFEGCFATTAAYQIEGAVKEGGRPSITDYYTCHLEPSTKNGANGVACDHHRYVED 65
DB      29 SFPKGFVFGTASSAFQHEGAVKAEGRGTTIDTSHFFGKITDSSNAVAADQYRYVED 88
OY      66 FDLITKYGAKAYRSLSMSRIIPLEGRLDPVNEGIEIEFYSLDALLRGGITPVWTLTYHW 125
DB      89 VQLKMGNGMDYRPSISWTRIFPMNG-VGHINENGIDHYNKLINALLAKGIEPYVTLTYHW 146
OY      126 DLPOLADHYCGMIVNEBVEQLDEFRYARLCEERGRDPQNMNITINXPVIOAIYGAATGSN 185
DB      147 DLPOLADHYRIGMLNPQIIN-DFAAYAEVCCQRGRDKVMMITTEBHTTFALIGYDVGLQ 205
OY      186 APGSSSINKKST-EGNTATPEWTLAKGAQIMSHARAFAVYSRDRFPSQGOIGISLNGDY 243
DB      186 APGSSSINKKST-EGNTATPEWTLAKGAQIMSHARAFAVYSRDRFPSQGOIGISLNGDY 243
  
```

Db 206 AFGRTIILFKLTREGNSSTEPYIVGHNVILTHATVSDIYRKVKAKOGSLGIAFDWM 265
QY 244 YEPMDNSNEPROKEAERMEHEHIGFANPIFKKDYPPSMKKOLGERIPALTPADPAILN 303
Db 266 FEP-BSNTEDELEAQRADDFOLGWFPLPLMF-GYFSSMSRVSRLPVTGSSGLVK 323
QY 304 AGETDFYGMNYYTSQFAHLLDGPVETDYLGAHHEHQNKDQ-----SPVGEESGL 354
Db 324 -GSIDFVGINHTTYARN-----NATWLIGTLHDAVSDGTYTLPPKGLSTIGDRASS 377
QY 355 AMLBSCPMPFRKHARVYGLYK-PYITTEGCGPCGGEENMTCEBAVDPFRIRYFDSHL 413
Db 378 IWLTVLPGRMSRLMNYIKGRYGNPVPVFTTEGMDPNISILSRKDALDKARIKYHNDYL 437
QY 414 DSISKAITODGVVYGFAMALLDNLLEWSDGVPREFGTYFTDY--TYLKRTPKS 466
Db 438 SSLQASIKEDGCNNGVTFWMSLLDNWEWAAGISSRFGLFYFDYRDLNLRKYPKDS 491

RESULT 3
AAG28766
ID AAG28766 standard; protein, 516 AA.
XX
AC AAG28766;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 34108.
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-014576P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.

PR	03-AUG-1999;	99US-0147038P
PR	04-AUG-1999;	99US-0147204P
PR	04-AUG-1999;	99US-0147302P
PR	05-AUG-1999;	99US-0147192P
PR	05-AUG-1999;	99US-0147260P
PR	06-AUG-1999;	99US-0147303P
PR	06-AUG-1999;	99US-0147416P
PR	06-AUG-1999;	99US-0147493P
PR	09-AUG-1999;	99US-0147831P
PR	10-AUG-1999;	99US-0148171P
PR	11-AUG-1999;	99US-0148349P
PR	12-AUG-1999;	99US-0148941P
PR	13-AUG-1999;	99US-0148565P
PR	13-AUG-1999;	99US-0148684P
PR	16-AUG-1999;	99US-0149368P
PR	23-AUG-1999;	99US-0149029P
PR	23-AUG-1999;	99US-0149930P
PR	17-AUG-1999;	99US-0149426P
PR	20-AUG-1999;	99US-0149722P
PR	20-AUG-1999;	99US-0149723P
PR	20-AUG-1999;	99US-0149529P
PR	27-AUG-1999;	99US-0150864P
PR	27-AUG-1999;	99US-0151066P
PR	27-AUG-1999;	99US-0151080P
PR	31-AUG-1999;	99US-0151303P
PR	31-AUG-1999;	99US-0151438P
PR	01-SEP-1999;	99US-0151930P
PR	07-SEP-1999;	99US-0152636P
PR	10-SEP-1999;	99US-0153070P
PR	13-SEP-1999;	99US-0153788P
PR	15-SEP-1999;	99US-0154039P
PR	16-SEP-1999;	99US-0154779P
PR	22-SEP-1999;	99US-0155139P
PR	23-SEP-1999;	99US-0155486P
PR	24-SEP-1999;	99US-0155659P
PR	28-SEP-1999;	99US-0156458P
PR	29-SEP-1999;	99US-0156566P
PR	04-OCT-1999;	99US-0157117P
PR	05-OCT-1999;	99US-0157573P
PR	06-OCT-1999;	99US-0157865P
PR	07-OCT-1999;	99US-0158023P
PR	08-OCT-1999;	99US-0158232P
PR	12-OCT-1999;	99US-0158369P
PR	13-OCT-1999;	99US-0159233P
PR	13-OCT-1999;	99US-0159294P
PR	13-OCT-1999;	99US-0159295P
PR	14-OCT-1999;	99US-0159329P
PR	14-OCT-1999;	99US-0159331P
PR	14-OCT-1999;	99US-0159333P
PR	14-OCT-1999;	99US-0159637P
PR	14-OCT-1999;	99US-0159638P
PR	18-OCT-1999;	99US-0159584P
PR	21-OCT-1999;	99US-0160741P
PR	21-OCT-1999;	99US-0160767P
PR	21-OCT-1999;	99US-0160768P
PR	21-OCT-1999;	99US-0160770P
PR	21-OCT-1999;	99US-0160814P
PR	21-OCT-1999;	99US-0160815P
PR	22-OCT-1999;	99US-0160980P
PR	22-OCT-1999;	99US-0160981P
PR	22-OCT-1999;	99US-0160989P
PR	25-OCT-1999;	99US-0161404P
PR	25-OCT-1999;	99US-0161405P
PR	25-OCT-1999;	99US-0161406P
PR	26-OCT-1999;	99US-0161359P
PR	26-OCT-1999;	99US-0161360P
PR	26-OCT-1999;	99US-0161361P
PR	28-OCT-1999;	99US-0161920P
PR	28-OCT-1999;	99US-0161992P

PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match	39.2%;	Score 1037; DB 3; Length 516;
Best Local Similarity	44.1%;	Pred. No. 2e-94;
Matches	209;	Conservative 85; Mismatches 156; Indels 24; Gaps 10
OY	6	ALPNDPEWGFATAAYQIEGAVKEGGRGSPISWDYTCHEPSPRTNGANGVACDHHRYED 65
DB	31	SFPKGFVFGTASSAFQHEGAVKAEGRGPTTIDWTSHTGKITTFESNADVAADQYHRYEED 90
OY	66	FDLLTKYAKAYRPSLSMSRIILPLGRDPNMEGIEFEYSKLIIDALLRGITPWTLYHW 125
DB	91	VQLKNNMGMDAYRPSISWTRIFPNG--VGHINEAGIDHYNKLINALAKGLEPVTLYHW 148
OY	126	DLPOLHRYGAMNVEBEVQDFFERYARLCEFERFGDRYQNMITIXKPMIOIYGYATGSN 185
DB	149	DLPOLHRYIGMLNPQIIN-DFAAYAEVCTQREGDRVKHMITTEPHTTFALQSDVGLQ 207
OY	186	APGSSSINKHST--EGNTATEPWLAKGAQIMSHARAVALVYSDRPPRSQKGOIGISLNDY 243
DB	208	APGRCTILFLKTCREGNSSTEPYIVGHNVILTHATVSDIYRKKYAKGGSGIAGFVWM 267
OY	244	YEPDSDNEPRKKAARERMEHIGFANPIFLKQDPSMSMKQCGERLPALTRADPALN 303
DB	268	FEP-ESNKTEDIEAAQRAQDQFLQMFDPPLMF-GDYPSSMSRVSRLPVFTGSSSLVK 325
OY	304	AGETDFEGMNYTOSOFARHLDPVPETDYLGAIHEHQNKG-----SPVGEESGL 354
DB	326	GSLDPFVGINHTYYARN-----NATNLIGTLHDAVSDSGYTYLPFKGLSTIGDRASS 379
OY	355	AMLRSCPDMFRKHLARVYGLYK-PIYITENGCPCGEEHNTCEBAVNDPPIRYFDSHL 413
DB	380	IWLIVPRGKMSLMNVIYIGRYGNPVPFITENGMDDPNSILSRQALDAKRIKYHNDYL 439
OY	414	DSISKAITQDGVVYKGYFPAALLDNLKESDQYGRFGYTFPDY-TTLKRTPKS 466
DB	440	SSLQASIKEDCNKYGYPVMSLIDNWEAAGYSRFGIYFVDYRDLKRYPKDS 493
RESULT 4		
AG828765		
ID	AG828765	standard; protein; 523 AA.
AC		
AG828765;		
XX		
DT	17-OCT-2000	(first entry)
XX		
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 34107.
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000;	2000EP-00301439.
XX		
PR	25-FEB-1999;	99US-0121825P.
PR	05-MAR-1999;	99US-0123180P.
PR	09-MAR-1999;	99US-0123548P.
PR	23-MAR-1999;	99US-0125788P.
PR	25-MAR-1999;	99US-0126264P.
PR	29-MAR-1999;	99US-0126785P.
PR	01-APR-1999;	99US-0127462P.
PR	06-APR-1999;	99US-0128234P.
PR	08-APR-1999;	99US-0128714P.
PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.

PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 26-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-014256P.
PR 14-MAY-1999; 99US-0144218P.
PR 14-MAY-1999; 99US-0144219P.
PR 14-MAY-1999; 99US-0144221P.
PR 14-MAY-1999; 99US-0144370P.
PR 18-MAY-1999; 99US-0144768P.
PR 19-MAY-1999; 99US-0144941P.
PR 20-MAY-1999; 99US-015124P.
PR 21-MAY-1999; 99US-0153533P.
PR 24-MAY-1999; 99US-0153629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-013911P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-014091P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145088P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.

AA029189
ID AA029189 standard; protein; 549 AA.
AC AA029189;
XX
XX 17-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 34687.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PM
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-00301439.
PF
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137502P.
PR 04-JUN-1999; 99US-0137722P.
PR 07-JUN-1999; 99US-0138094P.
PR 08-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0143542P.
PR 13-JUL-1999; 99US-0143624P.
PR 14-JUL-1999; 99US-0144005P.
PR 15-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145193P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.

PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159283P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 36.5%; Score 965; DB 3; Length 549;

Best Local Similarity 42.3%; Pred. No. 3,6e-87; Matches 204; Conservative 85; Mismatches 169; Indels 24; Gaps 13;

QY 6 ALPNDFEWGATAYQIEGAVKEGGRPSIMDTYCHLEPERS-TNGANGDVACDHRYHDE 64
DB 7 SFRKGFLEGRASSSYQEGAVNEGARGOSVMDHFSNFPRISSSGNVAVDFYHRYKE 66
QY 65 DFLLLTKYKAKAVRFSISMSRIIPUGRLDPVNEEGIEFSKLLDALLRGITPWTLYH 124
DB 67 DIRKMKOINNDSEFRLSIAMPRLVLPYGRKDRGVSEEGIKFYNDVDELLEITPLVTIFH 126
QY 125 WDLPOLHHDYGGMLNVEEYQOLDEPARYLCEFEFGDRVONWITINXPMIOAITGYATGS 164
DB 127 WDLIPQDLEDEYGGFLS-EQITIDDRDYASLCFEFGFGRVSLWCTWNPWYSVAGYDTGR 185
QY 185 NAGRRSS--INKISTEGNTATEPWLAKQAIMSHARAVAVYSRDFRPSQKGQIGISLNGD 242

DB 186 KAPGRCSKYVNGASVAGMSGYEAYIVSHNMLLAHAEAVEVF-RKCDHIRNGQIGIAHNP 244
QY 243 YYPEMDSNPRDDEAREMERMEFHIGFANPIFLKQYPSMKQGLERLPALEPAFAL 302
DB 245 WYEPYDSDDDVEGCNRRANDFMLGWHQHPYAC-GDYETMKSVQDRLPSPFPEOSKUL 303
QY 303 NAGETDFYGMNYYTSQFA--RHLDGPVP-ETDYLGAIHEHONKDGSPVGEESGLAWL 357
DB 304 -IGSCDYGINYSSLFVKSIKHVDPTQPTWRDQ--GVDMKKTNIIDGQIAKQGSSEWS 360
QY 358 RSCPDMPFRKLARVYGLYK-PYITENGCPCEG----ENMTCEAVNDPFRIRYFDS 411
DB 361 FYTPGTGLRNILKVKYKTYGNPILITENG---GEVAEQSQSLWYNPSIDTERLEYEG 417
QY 412 HLDISKAITQDGVVVKYFAMALLNLEMSDGYGRFGVTFPDYTT-LKRPFKSALV 470
DB 418 HIAIHQALHEDGVRVEGYVWSLLDNFENSSGYGVYGLYIDYKQGLRRYKMSALWL 477
QY 471 KD 472
DB 478 KE 479

RESULT 7
ID AAG29188 standard; protein; 560 AA.

AC AAG29188;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 34686.

KM Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PE 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134263P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139761P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 19-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145129P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.

PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147202P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148655P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 25-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152353P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 13-OCT-1999; 99US-0159325P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160816P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.

PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161932P.
 PR 28-OCT-1999; 99US-0161933P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 36.5%; Score 965; DB 3; Length 560;
 Best Local Similarity 42.3%; Pred. No. 3.7e-87;

Matches 204; Conservative 85; Mismatches 169; Indels 24; Gaps 13;

QY 6 ALPNDEWGFATAYOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYHRDE 64
 DB 18 SFKGLFTGASSSYQYEGAVNEGARGQSWDFHSNRPFRIRISDSDGNVAVDYFHYRKE 77
 QY 65 DFDLTKYGAAYRSLMSRIIPLCGRDPVNEEGIEFYSKIDALLRRGITPWTLYH 124
 DB 78 DIRMMDINMDSFRLSIAMPRLVPGKRDGVSEEGIKFYNDVIDELANEITPLVTIFH 137
 QY 125 WDLPOALHRYGGMVNEEVOUDFERYARLCFERFGDRVONMTTIXKPMQATGYATGS 184
 DB 138 WDIPOQLEDEYGGFLS-EQIIDDYFVASCERFGRVSLMCTMNEPMWYSVAGYDTGR 196
 QY 185 NAGGRSS--INKSTEGNTATEPMLAGKAOIMSHARAVYSRDFPSQKQIGISLNGD 242
 DB 197 KAGRGSKYVNGASVAGMSGEYAVIYSHNMLLAHAVEVF-RKCHIKKQIGIAHNPL 255
 QY 243 YEPWDSNEPRDEKAERMEFIHGFANPIFLKKDYPESMKKQIGERLPALTPADFAIL 302
 DB 256 WYEPYPSDPDVEGCNRAMDFMLGWHQHPAC-GDYPETMKKSVDRLPSFTPEQSKKL 314
 QY 303 NAEETPFYGNMYTSQFA---RHLDGVP--ETDYLGAIHEHENDGSGVGESEGLAWL 357
 DB 315 -IGSCDYVGINYSSLFVKSIRKVDPTQPTWRDQ--GVMWMTNIDGKQIAKQSGSEWS 371
 QY 358 RSCPDWFRKHLARVYGLYK-PYITENGCPCE-----ENMTCEAVNDPFRIRYFDS 411
 DB 372 FTYPTGLRNILKVKKTYGNPILITNGY---GEVAEQSOSLYMNPISIDTERLEYIEG 428
 QY 412 HLDISKAITQDGVVVKGYFAMALLDNLEWSDGYGPRFGVTFDYTT-LKRTPKKSALVL 470
 DB 429 HIAHIOAHIEDGVREGVYVMSLIDNFEMNSGYGVYGLYIDYKDLRRYPMKSALWL 488
 QY 471 KD 472
 DB 489 KE 490

RESULT 8

ID ADC03121 standard; protein; 560 AA.

AC ADC03121;

DT 18-DEC-2003 (first entry)

XX A. thaliana family 1 glucoside hydrolase PEN2.

KM beta-glucosidase; antibacterial; fungicide; plant protection; plant;
 KW PEN2; family 1 glycoside hydrolase.

OS Arabidopsis thaliana.

PN W02003074688-A2.

PD 12-SEP-2003.

PF 06-MAR-2003; 2003MO-EP002315.

PR 06-MAR-2002; 2002EP-00004400.

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Iapka V, Scheel D, Schulze-Lefert P, Rosahl S, Landtag J;
 PI WPI; 2003-712890/67.
 XX N-PSDB; ADC03120.
 DR New polynucleotide encoding a beta glucosidase plant protein, useful for
 XX the preparation of a plant protection composition, and/or for
 PT establishing or enhancing pathogen resistance in plants.
 PS Claim 1; SEQ ID NO 2, 89pp; English.
 CC The invention relates to a novel polynucleotide encoding a beta-
 CC glucosidase protein. A polypeptide of the invention has antibacterial and
 CC fungicide activity. The identified compound or hydrolysis product of the
 CC invention is useful for preparing a plant protection composition. The
 CC present sequence represents the Arabidopsis thaliana PEN2 protein.
 SQ Sequence 560 AA;

Query Match 36.5%; Score 965; DB 7; Length 560;

Best Local Similarity 42.3%; Pred. No. 3.7e-87;

Matches 204; Conservative 85; Mismatches 169; Indels 24; Gaps 13;

QY 6 ALPNDEWGFATAYOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYHRDE 64
 DB 18 SFKGLFTGASSSYQYEGAVNEGARGQSWDFHSNRPFRIRISDSDGNVAVDYFHYRKE 77
 QY 65 DFDLTKYGAAYRSLMSRIIPLCGRDPVNEEGIEFYSKIDALLRRGITPWTLYH 124
 DB 78 DIRMMDINMDSFRLSIAMPRLVPGKRDGVSEEGIKFYNDVIDELANEITPLVTIFH 137
 QY 125 WDLPOALHRYGGMVNEEVOUDFERYARLCFERFGDRVONMTTIXKPMQATGYATGS 184
 DB 138 WDIPOQLEDEYGGFLS-EQIIDDYFVASCERFGRVSLMCTMNEPMWYSVAGYDTGR 196
 QY 185 NAGGRSS--INKSTEGNTATEPMLAGKAOIMSHARAVYSRDFPSQKQIGISLNGD 242
 DB 197 KAGRGSKYVNGASVAGMSGEYAVIYSHNMLLAHAVEVF-RKCHIKKQIGIAHNPL 255
 QY 243 YEPWDSNEPRDEKAERMEFIHGFANPIFLKKDYPESMKKQIGERLPALTPADFAIL 302
 DB 256 WYEPYPSDPDVEGCNRAMDFMLGWHQHPAC-GDYPETMKKSVDRLPSFTPEQSKKL 314
 QY 303 NAEETPFYGNMYTSQFA---RHLDGVP--ETDYLGAIHEHENDGSGVGESEGLAWL 357
 DB 315 -IGSCDYVGINYSSLFVKSIRKVDPTQPTWRDQ--GVMWMTNIDGKQIAKQSGSEWS 371
 QY 358 RSCPDWFRKHLARVYGLYK-PYITENGCPCE-----ENMTCEAVNDPFRIRYFDS 411
 DB 372 FTYPTGLRNILKVKKTYGNPILITNGY---GEVAEQSOSLYMNPISIDTERLEYIEG 428
 QY 412 HLDISKAITQDGVVVKGYFAMALLDNLEWSDGYGPRFGVTFDYTT-LKRTPKKSALVL 470
 DB 429 HIAHIOAHIEDGVREGVYVMSLIDNFEMNSGYGVYGLYIDYKDLRRYPMKSALWL 488
 QY 471 KD 472
 DB 489 KE 490

RESULT 9

ID ABP71306 standard; protein; 545 AA.

AC ABP71306;

DT 28-APR-2003 (first entry)

DE Prunus serotina prunasin hydrolase.

XX Prunasin hydrolase; transcription; vascular tissue; gene expression;
 KW abiotic stress; biotic stress.

PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142380P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144085P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144325P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144352P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 21-JUL-1999; 99US-0144884P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145085P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 22-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151030P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0156559P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159638P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161559P.
 PR 26-OCT-1999; 99US-0161609P.
 PR 26-OCT-1999; 99US-0161610P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161933P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 35.4%; Score 939; DB 3; Length 512;
 Best Local Similarity 41.4%; Pred. No. 1,3e-84;
 Matches 197; Conservative 84; Mismatches 175; Indels 20; Gaps 11;
 QY 6 ALPNDFEMGFATAYOIEGAVKGGGPGPSIWDTYCHLBPRTNGANGDVACDHYHYKDED 65
 DB 43 SFPKGFVFATATAYOVEGTHODGRGPGSIWDAFVKI PKIAXNAATAEITVDQYHRYKED 102
 QY 66 FDLTLTYGAKAYRFSWSRRIPLGGRDPVNEEGIEFYSKILDALRGGITPMTLYHW 125
 DB 103 VDLAKKLNDPARYFSSISWSRIPEGS--GKNMKGVATYRNLIDYVQKITYAALYHY 160
 QY 126 DLFOALHRYGGLNVAEEVQDPERYARLCFERFGDRVQNMWITINXPMWIOAIYVATGSN 185

Db 161 DLPLALENKKYKLLG-RQVYKDFADYAEFCYKTFGDRVKNNMTFNEPRVVALGYDNGIF 219
Qy 186 AGR-SINSHSTEGNTATEPMLAGKAQIMSHARAVASRFRSQKQIGISLNGDY 244
Db 220 AGRCKKAGCTNCTEGNSATEPPIVTHHLILALAAVORRKYKYOAKQKRGVGLIDFWY 279
Qy 245 EPMDSNEPPDKAEARERMEFHIGWPNPIFLKKDYPSKCOLGRPLATPADPAIINA 304
Db 280 EPL-TRSKADNLAAQRAFDHFHIGWFIHPL-VYGEVPTKQNIIVKERLRFTEKEVKNK- 336
Qy 305 GETDFYGMVYTSQFARHLDGVPPEPDYIGALHENO-----KDGSPVEESGLAMR 358
Db 337 GSIDFPGINQVYTTY---NSEPHPTTKPKDGLYQODMNVFEGFALGKPIGRAYSSWLY 393
Qy 359 SCPDMFRKILARVYGLYKRP-IYITENGCPCEENMTCEBAVNDPFRIRYFDSHLSIS 417
Db 394 NVPWGKRYKALMYMKERYGNPTWILSENGWDDPG--NVTLAOGHLHTTIKIKYKYLTIUK 451
Qy 418 KAITDDGVVVKGYFPMALIDNLEMSDGYGPRGVTFTDTYTIKRTPKKSALVLKDM 473
Db 452 KA-RDDGANVVGYPFAMSLIDNEMLSGYTSRFGIYVVDYKTLKRYPKMSAQWFKQL 506
RESULT 11
AAG14174
ID AAG14174 standard; protein, 521 AA.
AC AAG14174;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 13936.
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0138847P.
PR 16-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.

PR	08.-JUN.-1999;	99US-01455913;
PR	02.-AUG.-1999;	99US-0146386P;
PR	02.-AUG.-1999;	99US-0146388P;
PR	02.-AUG.-1999;	99US-0146389P;
PR	03.-AUG.-1999;	99US-0147038P;
PR	04.-AUG.-1999;	99US-0147204P;
PR	04.-AUG.-1999;	99US-0147202P;
PR	05.-AUG.-1999;	99US-0147192P;
PR	05.-AUG.-1999;	99US-0147260P;
PR	06.-AUG.-1999;	99US-0147476P;
PR	06.-AUG.-1999;	99US-0147493P;
PR	09.-AUG.-1999;	99US-0147935P;
PR	10.-AUG.-1999;	99US-0148117P;
PR	11.-AUG.-1999;	99US-0148319P;
PR	12.-AUG.-1999;	99US-0148354P;
PR	13.-AUG.-1999;	99US-0148664P;
PR	13.-AUG.-1999;	99US-0149368P;
PR	17.-AUG.-1999;	99US-0149175P;
PR	18.-AUG.-1999;	99US-0149466P;
PR	20.-AUG.-1999;	99US-0149722P;
PR	26.-AUG.-1999;	99US-0150884P;
PR	27.-AUG.-1999;	99US-0151065P;
PR	27.-AUG.-1999;	99US-0151066P;
PR	30.-AUG.-1999;	99US-0151080P;
PR	31.-AUG.-1999;	99US-0151030P;
PR	01.-SEP.-1999;	99US-0151438P;
PR	01.-SEP.-1999;	99US-0151930P;
PR	07.-SEP.-1999;	99US-0152363P;
PR	10.-SEP.-1999;	99US-0153707P;
PR	13.-SEP.-1999;	99US-0153758P;
PR	15.-SEP.-1999;	99US-0154018P;
PR	16.-SEP.-1999;	99US-0154032P;
PR	20.-SEP.-1999;	99US-0154779P;
PR	22.-SEP.-1999;	99US-0155139P;
PR	23.-SEP.-1999;	99US-0155486P;
PR	24.-SEP.-1999;	99US-0155659P;
PR	28.-SEP.-1999;	99US-0156458P;
PR	29.-SEP.-1999;	99US-0156569P;
PR	04.-OCT.-1999;	99US-0157117P;
PR	05.-OCT.-1999;	99US-0157753P;
PR	06.-OCT.-1999;	99US-0157865P;
PR	07.-OCT.-1999;	99US-0158029P;
PR	08.-OCT.-1999;	99US-0158232P;
PR	12.-OCT.-1999;	99US-0158366P;
PR	13.-OCT.-1999;	99US-0159283P;
PR	13.-OCT.-1999;	99US-0159284P;
PR	13.-OCT.-1999;	99US-0159285P;
PR	14.-OCT.-1999;	99US-0159330P;
PR	14.-OCT.-1999;	99US-0159332P;
PR	14.-OCT.-1999;	99US-0159593P;
PR	14.-OCT.-1999;	99US-0159633P;
PR	14.-OCT.-1999;	99US-0159637P;
PR	14.-OCT.-1999;	99US-0159638P;
PR	18.-OCT.-1999;	99US-0159584P;
PR	18.-OCT.-1999;	99US-0160741P;
PR	21.-OCT.-1999;	99US-0160767P;
PR	21.-OCT.-1999;	99US-0160768P;
PR	21.-OCT.-1999;	99US-0160774P;
PR	21.-OCT.-1999;	99US-0160781P;
PR	22.-OCT.-1999;	99US-0160989P;
PR	22.-OCT.-1999;	99US-0161404P;
PR	25.-OCT.-1999;	99US-0161405P;
PR	25.-OCT.-1999;	99US-0161406P;
PR	26.-OCT.-1999;	99US-0161359P;

PR 26-OCT-1999; 99US-0161360P
PR 26-OCT-1999; -99US-0161361P
PR 28-OCT-1999; 99US-0161920P
PR 28-OCT-1999; -99US-0161992P
PR 28-OCT-1999; 99US-0161993P
PR 29-OCT-1999; 99US-0162142P

Query Match	35.5%	Score 939	DB 3	Length 521
Best Local Similarity	41.4%	Pred. No.	1.3e-84	
Matches 197	Conservative	84	Mismatches	175
			Indels	20
			Gaps	11

```
Qy      6 ALPNDEWGFATAYQIEGAVKEGGRGESIMDITYCHLEPSRTNGANGDVACDHYHRXDED 655
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      52 SFPKGFVFGTATSAYQVEGETHODGRGSIWDAEVIKPKIAKNTAEITVDQYHRKED 111
```

```
Qy      66 FDLTKYGAAAYRFSLSWSRIIPLGRLDPVNEEGIEFYSKIIDLARRGITWVTLYHW 122  
        |||::|||::|||::|||::|||::|||:  
Db     112 VDLMKLTNPDAYRFSSISMSRFPFGS--GKNVMKGVAYYNRLLDYMVGKITPYANLHYH 165
```

```

Qy      126 DLPQALHDRYGGMLNVEEVLDFERARLCEERFGDRVQNMWTTINXPMWIAIYGATSSN 185
      ||| ||| :| :| ||| ||| :| :| ||| ||| :| :| ||| ||| :| :| |||
Db      170 DLPALLENKTKGLLG-RQVVNDPADVAEFCYKTFGDRVKNMMTFENEPVVAAALGYDNGIF 228

```

```
Oy      186 APCR-SSINKHTEGNTATEPWLAKAQIMSHARAVAVSRDFRPSQXQIGISLNGDY 24  
        ||||| : ||||| : ||||| : ||||| : ||||| :  
Db      229 APRGRSKAFSGCTEGNSATEEYIVTHLLLAHAADAQRKRYQAQKGRAVGLLDFTWY 28
```

```
Oy      245 EPMDSNPEPKKEAERMEFHIGWPNPIFLKKDYPSMKKOLGERLPALTPADFAILNA 304  
        || :||| :|||| : : :||| :||| : :  
Db      289 EPL-TRSKADNLAAQRARDFIHGWFHPL-VYGEYPTKMONTVKERLPEKTEKEVKNMY- 347
```

Qy 305 GETPEYGMNYYTSQPARHLDGPPVETIDYLGAIHENQEN-----KDSSPVGEESGLAMLR 356
|||::|||::|||::|||::|||
Db 346 GSIDEVGINGYTTY--MSEPHPTTKPKDLGQQDNNVFEGFAKLGRIGPRAYSSWLY 407

Qy 359 SCPDMFRKRLARVYLGYGPK-IYITENGCCPPGEEANNTCCEAVNDPRIRYPFSLDLS 417

Db 403 NVPMGMYYKALMYMKERYGNPTMILSNGMDPG-NVTLAGLHDTTRIKKYDYLTNIK 466

Qy . 418 KAITQDGVVVKGYFAMALLDNLNLENSDGYGPRFGVFTDYTLTKRTPKKSALVLKDM 417

Db 461 KA-RDDGANVGVFAMSLLDNFEMLSGYSRFGIYVDYDKTLKRYPKMSAQWFQJL 515

RESULT 12
ABP99404

XX
AC
XX

ABP99404;

XX
DE Arabidopsis thaliana polypeptide SEQ ID NO 93.
XX

KW herbicide-resistant.
XX
OS *Arabidopsis thaliana*.

PN	W0200266660-A2
XX	
PD	29-AUG-2002.

PF	13-FEB-2002; 2002WO-EP001466
XX	
PR	16-FEB-2001; 2001DE-01007843

XX (META-) METANOMICS GMBH & CO KGAA
PA
XX

XX
DR WPI; 2002-674953/72.

PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 03-JUL-1999; 99US-0142055P.
PR 04-JUL-1999; 99US-0142380P.
PR 05-JUL-1999; 99US-0142803P.
PR 06-JUL-1999; 99US-0142920P.
PR 07-JUL-1999; 99US-0142977P.
PR 08-JUL-1999; 99US-0143542P.
PR 09-JUL-1999; 99US-0143624P.
PR 10-JUL-1999; 99US-0143624P.
PR 11-JUL-1999; 99US-0143624P.
PR 12-JUL-1999; 99US-0143624P.
PR 13-JUL-1999; 99US-0143624P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0143624P.
PR 16-JUL-1999; 99US-0143624P.
PR 17-JUL-1999; 99US-0143624P.
PR 18-JUL-1999; 99US-0143624P.
PR 19-JUL-1999; 99US-0143624P.
PR 20-JUL-1999; 99US-0143624P.
PR 21-JUL-1999; 99US-0143624P.
PR 22-JUL-1999; 99US-0143624P.
PR 23-JUL-1999; 99US-0143624P.
PR 24-JUL-1999; 99US-0143624P.
PR 25-JUL-1999; 99US-0143624P.
PR 26-JUL-1999; 99US-0143624P.
PR 27-JUL-1999; 99US-0143624P.
PR 28-JUL-1999; 99US-0143624P.
PR 29-JUL-1999; 99US-0143624P.
PR 30-JUL-1999; 99US-0143624P.
PR 31-AUG-1999; 99US-0143624P.
PR 01-SEP-1999; 99US-0143624P.
PR 02-SEP-1999; 99US-0143624P.
PR 03-SEP-1999; 99US-0143624P.
PR 04-SEP-1999; 99US-0143624P.
PR 05-SEP-1999; 99US-0143624P.
PR 06-SEP-1999; 99US-0143624P.
PR 07-SEP-1999; 99US-0143624P.

PR 10-SEP-1999; 99US-0153070P.
PR 11-SEP-1999; 99US-0153758P.
PR 12-SEP-1999; 99US-0154018P.
PR 13-SEP-1999; 99US-0154039P.
PR 14-SEP-1999; 99US-0154039P.
PR 15-SEP-1999; 99US-0154039P.
PR 16-SEP-1999; 99US-0154039P.
PR 17-SEP-1999; 99US-0154039P.
PR 18-SEP-1999; 99US-0154039P.
PR 19-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154039P.
PR 21-SEP-1999; 99US-0154039P.
PR 22-SEP-1999; 99US-0154039P.
PR 23-SEP-1999; 99US-0154039P.
PR 24-SEP-1999; 99US-0154039P.
PR 25-SEP-1999; 99US-0154039P.
PR 26-SEP-1999; 99US-0154039P.
PR 27-SEP-1999; 99US-0154039P.
PR 28-SEP-1999; 99US-0154039P.
PR 29-SEP-1999; 99US-0154039P.
PR 30-SEP-1999; 99US-0154039P.
PR 01-OCT-1999; 99US-0154039P.
PR 02-OCT-1999; 99US-0154039P.
PR 03-OCT-1999; 99US-0154039P.
PR 04-OCT-1999; 99US-0154039P.
PR 05-OCT-1999; 99US-0154039P.
PR 06-OCT-1999; 99US-0154039P.
PR 07-OCT-1999; 99US-0154039P.
PR 08-OCT-1999; 99US-0154039P.
PR 09-OCT-1999; 99US-0154039P.
PR 10-OCT-1999; 99US-0154039P.
PR 11-OCT-1999; 99US-0154039P.
PR 12-OCT-1999; 99US-0154039P.
PR 13-OCT-1999; 99US-0154039P.
PR 14-OCT-1999; 99US-0154039P.
PR 15-OCT-1999; 99US-0154039P.
PR 16-OCT-1999; 99US-0154039P.
PR 17-OCT-1999; 99US-0154039P.
PR 18-OCT-1999; 99US-0154039P.
PR 19-OCT-1999; 99US-0154039P.
PR 20-OCT-1999; 99US-0154039P.
PR 21-OCT-1999; 99US-0154039P.
PR 22-OCT-1999; 99US-0154039P.
PR 23-OCT-1999; 99US-0154039P.
PR 24-OCT-1999; 99US-0154039P.
PR 25-OCT-1999; 99US-0154039P.
PR 26-OCT-1999; 99US-0154039P.
PR 27-OCT-1999; 99US-0154039P.
PR 28-OCT-1999; 99US-0154039P.
PR 29-OCT-1999; 99US-0154039P.

Query Match 35.0%; Score 925.5; DB 3; Length 506;
Best Local Similarity 43.1%; Pred. No. 2, 9e-83;
Matches 204; Conservative 71; Mismatches 169; Indels 29; Gaps 14;

QY 8 PNDPFGATTAAYQIGAVVGGGSGPSIMDTYCHLEPSR-TNGANGDVACDHYRYDEDF 66
DB 37 PEDFIFGSATSAVQVGGAGHEDGRGSIIMDTSEKPEKPKKDSNGSVADNSHLYKEDV 96
QY 67 DLLTKYGAAYRPSLSWSRIIPVGRGLDPVNEBGEIEFYSKLDALLRGITTPVTLVHMD 126
DB 97 ALHQIGFNAVYRISISMSRLIPRGNLKGGINQIDYNNLIVELLISKIKPRATMFMHD 156
QY 127 LPQALHRYGGMVNEVQDLPERYARLCPERGGDRVQWITINXWICAIYIYAGSNA 186
DB 157 TPQALBDAYGGFGAGAIIV-DFRDYADICFKNGEDVKMKMTINEPLTVVQOQYVAGVMA 215
QY 187 PGRSS--INKSTEGTATEPMLAGKAQIMSHARAVYSRDRPSQOIGISLNDGY 244
DB 216 PGRCSKFTNPNCTDNGATTEPIYVGNLILSHAAVQVIREKTKAQQOQVGIALNAGWN 275
QY 245 EPWDSNEPRDKAEARMEFHIGWFPANPIFLKKDYESNKKQIGERLPAITPADFAILNA 304
DB 276 LPY-TESPKDRILAAABAMAFTFDYFMEPIYVTK-YVDMVNNVKGRLPIFTAQOSKWLK- 332
QY 305 GETDPFGANYTTSQPARHLDGPVETD---YLGAIEHQNKGDSVGBESSGLAMLRSC 361
DB 333 GSYDFLIGINYSSTYAK--DVPCSTYDVTWTFSDPCASVTGERDGVIPGKAASDWLLIYP 390

```
QY 362 DMFRKHLARVYGLYCK-----PI-YITENGCPGCEENMTCEAVNDPFRIRFPDSDS 415
DB 391 KGIROLV-----LVAKYKEXPVMTITENG-----RDESTNKIFLKQDGRIDYAKRHEM 441
QY 416 ISKAITOGVVVKGYPAMALLDNLLEWSDGYGPRFGVTFTDYTT-LKRTPKKSA 467
DB 442 VODAISV-GANVKGFFAMSLDNLNFEWAMGYTVRFGLVYVDFKQCKRYPKSA 493

RESULT 14
ID AAY52193 standard; protein; 657 AA.
XX
AC AAY52193;
XX
DT 06-AUG-2003 (revised)
DT 15-FEB-2000 (first entry)
XX
DE Beta glucosidase (Bgl A) amino acid sequence.
XX
KM Bgl A; beta glucosidase; cellulose; cellulose biomass.
XX
OS Orpinomyces sp. PC-2.
XX
FH Key Location/Qualifiers
FT Peptide 1..16
FT /note= "Putative signal peptide"
FT Protein 17..641
FT /note= "Beta-glucosidase"
XX
PN WO9957287-A2.
XX
PD 11-NOV-1999.
XX
PF 06-MAY-1999; 99WO-US010106.
XX
PR 06-MAY-1998; 98US-0084494P.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Li X, Ljungdahl LG, Chen H, Ximenes EA;
XX
XX WPI; 2000-023580/02.
DR N-PSDB; AA237865.
XX
XX New orpinomyces PC-2 beta-glucosidase coding sequences and protein.
XX
PS Claim 1; Page 35-36; 63pp; English.
XX
CC This is the beta-glucosidase (BglA) amino acid sequence from Orpinomyces
CC sp. strain PC-2. Beta-glucosidase acts to liberate D-glucose units from
CC cellobiose, cello-oligosaccharides and other glucosides. A host cell
CC containing the BglA DNA sequence can be used to produce a beta-
CC glucosidase in a host cell other than Orpinomyces sp strain PC-2. The
CC fungal beta-glucosidase is useful for the improved saccharification of
CC cellulosic materials. Effective utilization of cellulosic biomass through
CC biological processes is one approach to overcoming the shortage of food,
CC feeds and fuels, expected as a consequence of the explosive increase in
CC human population. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 657 AA;

Query Match 34.9%; Score 923.5; DB 3; Length 657;
Best Local Similarity 36.0%; Pred. No. 6.9e-83;
Matches 209; Conservative 82; Mismatches 158; Indels 131; Gaps 17;

QY 7 LPNDFFMGPAITAAYOIGAVVEGGRPSIMDTYCHLEP-----SRTGANGDVACDH 58
DB 74 LPADFCKMGAITAAAYQVGAWDEBGRGSWDHFTHLTPKVVESGDRSKDSTNGNACDS 133
QY 59 YHRVDEDFDLTKYGAAYRFFSLWSRIIPAG-----GLDPVNEGIEFFYSKLIDAL 112
DB 134 YHKRDEDEVKMLKMNAAKYRFSISWPRLLFPDQAKRKDGKMN-VNKGAEYIMVINTLL 192
```

```
QY 113 RRGITPVYTHYHNDLPOLADHRYGWLANEEVOLDERRVRLCPERRGDRVONMTITNXP 172
DB 193 KNDIVPFTLYHHWDLFPALHEIKYGMWDYHS-ODDPKAYAEFCFERRFGDRVKNMTITINEP 251
QY 173 MIOAIYGVATG-----SNAP-----GRSINGHSTEGCNTATPEPLAGKA 211
DB 252 WNVGVSGYRLGPGKAPRCTGGEAPRKLQNSTDLDEGGCSYEIGPYOSKNSEPLPANRV 311
QY 212 Q-----IMSHRAVAVYSRDPFSPQKQIGISLNGDYEPWDSNEPRDXE--- 256
DB 312 POKLEDVWCSHNILLGHAKAVKYRREKFKQKGLIGITVDGBAQIPWEVPGMTKKEYEN 371
QY 257 ---AERRMEFHGFWANPIFLKKDYBESWKQOLGERLPALTADPAIINAGSTDVYGNM 313
DB 372 NLKYANLAAEFRLGWSYSDPPMV-GDYPKSVKEMGDLDEFTEEKKILKSSSDFLGNM 430
QY 314 YYTSQF---ARHLDGVPEPTDYLGAIHEQENKDS-----PYG 349
DB 431 TYTAHMAAQAKNEDG-----SYIQPTAEANFPDNKXKMDNCKGRDGTCTIPTLG 485
QY 350 EESGLAM-----LNSCPDMPKHLARVYGLYKPIYITENGCPGCG----- 390
DB 486 SQAGSSWNTKFAPTIRIVGLNWFESK---RYEGLIKNGIVITENGCAQPNYKVARANDEVTK 542
QY 391 -----EENMTCEE-----AVNDPFRIRFPDSDHSISKAITODGVVNG 429
DB 543 KYFESTGQPKYADTYVEEDIEDNLEGLMHDYTIIDWYDYLKRLAYAVANDIVRG 602
QY 430 YFAMALLDNLLEWSDGYGPRFGVTFTDY---TLKRTPKKS 466
DB 603 YMAWSLLDNFEWENGYETRFQMTYIDFYNDKEMKRVPKDS 642

RESULT 15
AAG50272
ID AAG50272 standard; protein; 507 AA.
XX
AC AAG50272;
XX
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63689.
DE
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
PF
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
```


PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140685P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.

PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145195P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154799P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.

PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161320P.
PR	28-OCT-1999;	99US-0161932P.
PR	28-OCT-1999;	99US-0161933P.
PR	29-OCT-1999;	99US-0162142P.

Query Match	34.9%	Score 923;	DB 3;	Length 507;
Best Local Similarity	41.9%;	Pred. No. 5.2e-83;		
Matches 205; Conservative	75;	Mismatches 173;	Indels 36;	Gaps 16;

QY	8	PNDPEMFATPAAOIEAIVEGGRGSIINDYTHLEPSR--TNGANGDVAOCHRYHVEDF	66
Db	37	PKPITFBAALISAQVEGAHEDRGSIINDTSEKPEKIKDGTGSIASDSYHLYEDV	96
QY	67	DLTKYAKAKYRPSLSWRIIPGRLDPVNEEGIEFYSKIDALLRGRITPWYTLHYMD	126
Db	97	GLHLOIQFGFYRPSISWSRIIPRGMLKGGINOAQIDYNNININELLSKGIKPFATIFHMD	156
QY	127	LPQALHRYSGMNLVEVOLDPERYARLCFEEFGRGVQNMWLTINXPMWQALYGAITSNA	186
Db	157	TPQSLLEAYGCFEGAEIYN--DFEDYADICQKPFGRVGHMMTLEPLTVVQOQVAAWMA	215
QY	187	PGRSS--INKGSTGNATATEPMLAKQAOIMSHARAVAAYSRDFRPSQKQIGISLNDY	244
Db	216	PGGCSKFTYNNCTAGCATETPYVGHNTLILHGAEVKVKYREKYASQKQGVGINLNGMN	275
QY	245	EPDMSNEPRKEAERMEFHIGWFPANPIELKQOYBESMKQOLG--RLPALTPADFAILN	303
Db	276	LPIY--TESAEDRLNAAPMAAFPTYDFMEPLVYTGK--PYVMVNVNVDKGRIPPTAKOSIMLK	333
QY	304	AGETDFYGMNYTISQFARHLDPVDE-----TDVIGAIIHGHQNKDGSPIVGEESGLATL	357
Db	334	--GSYDFEIGIINYSSYAK--DVPCSENVTLFSDPCASVTBERE---GVPIGRPLAADWL	387
QY	358	RSCPDMFERKHLARVYGLYTK-----PI-YITENGCPCCGEENMTCEEAIVNDPERIRYEDS	411
Db	388	LIYFKGIRDLL-----LVAKYFKDPWMTYITENG----RDEASGKIDLKDSEKIDVYAQ	438
QY	412	HLDSIKAITODGVVVGKYFAWALLDNLEWSDGCVPRGCVFTPYTT--LKRTPKKSATLV	470
Db	439	HLKMWODALS--IGANVKGFAMSLIDNFMATGYSVTRGLVYVDPNDGRKRYPKKSAXNF	497
QY	471	KDMFAAROR	479
Db	498	RKLSEKR	506

Search completed: March 25, 2004, 15:55:08
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2004, 15:54:02 ; Search time 22 Seconds
(without alignments) 1135.771 Million cell updates/sec

Title: US-10-026-140-2

Perfect score: 2644

Sequence: 1 MESLALPDEFWGFATAY.....KSALVLKDFARQVKA 484

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Issued Patents AA:*

1: /cgn2_6/ptodata/2/1aa/5A COMB.pep:*

2: /cgn2_6/ptodata/2/1aa/5B COMB.pep:*

3: /cgn2_6/ptodata/2/1aa/6A COMB.pep:*

4: /cgn2_6/ptodata/2/1aa/6B COMB.pep:*

5: /cgn2_6/ptodata/2/1aa/PCFUS COMB.pep:*

6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	983	37.2	551	3 US-08-688-988-37	Sequence 37, Appl
2	981.5	37.1	507	4 US-09-914-841A-15	Sequence 15, Appl
3	976	36.9	509	3 US-08-688-988-40	Sequence 40, Appl
4	964.5	36.5	476	3 US-09-306-593-9	Sequence 9, Appl
5	924.5	35.0	531	3 US-08-688-988-39	Sequence 39, Appl
6	923.5	34.9	657	3 US-09-306-593-2	Sequence 2, Appl
7	910	34.4	513	2 US-09-122-230-7	Sequence 7, Appl
8	902	34.1	563	4 US-09-431-470-4	Sequence 4, Appl
9	896	33.9	531	3 US-08-688-988-38	Sequence 38, Appl
10	880.5	33.3	459	4 US-08-914-375C-72	Sequence 72, Appl
11	879	33.2	566	4 US-09-431-470-2	Sequence 2, Appl
12	863.5	32.7	524	3 US-08-688-988-34	Sequence 34, Appl
13	861	32.6	429	4 US-09-914-841A-1	Sequence 1, Appl
14	858.5	32.5	476	4 US-08-914-375C-69	Sequence 69, Appl
15	858.5	32.5	548	3 US-08-688-988-41	Sequence 41, Appl
16	857.5	32.4	470	4 US-08-914-375C-73	Sequence 73, Appl
17	856	32.4	528	3 US-08-688-988-6	Sequence 6, Appl
18	854	32.3	539	3 US-08-688-988-42	Sequence 42, Appl
19	849.5	32.1	540	3 US-08-688-988-43	Sequence 43, Appl
20	845	32.0	473	4 US-08-914-375C-71	Sequence 71, Appl
21	842	31.8	471	4 US-08-914-375C-70	Sequence 70, Appl
22	840.5	31.8	450	3 US-09-306-593-13	Sequence 13, Appl
23	839	31.7	446	3 US-09-306-593-11	Sequence 11, Appl
24	831	31.4	525	3 US-08-688-988-10	Sequence 10, Appl
25	826	31.2	438	4 US-09-134-078-16	Sequence 16, Appl
26	810	30.6	514	3 US-08-688-988-35	Sequence 35, Appl
27	807	30.5	469	3 US-09-306-593-8	Sequence 8, Appl

28	802	30.3	441	4 US-08-914-375C-66	Sequence 66, Appl
29	802	30.3	449	3 US-09-306-593-10	Sequence 10, Appl
30	791	29.9	439	4 US-08-914-375C-67	Sequence 67, Appl
31	790	29.9	425	3 US-08-688-988-36	Sequence 36, Appl
32	789	29.8	437	4 US-08-914-375C-65	Sequence 65, Appl
33	762	28.8	449	4 US-08-914-375C-68	Sequence 68, Appl
34	749	28.3	448	3 US-09-306-593-12	Sequence 12, Appl
35	747	28.3	441	4 US-08-914-375C-63	Sequence 63, Appl
36	745.5	28.2	356	4 US-08-914-375C-74	Sequence 74, Appl
37	726.5	27.5	450	4 US-08-914-375C-64	Sequence 64, Appl
38	718	27.2	1014	4 US-09-344-510B-3	Sequence 3, Appl
39	714	27.0	550	4 US-09-344-510B-4	Sequence 4, Appl
40	710	26.9	549	4 US-09-344-510B-2	Sequence 2, Appl
41	710	26.9	1012	4 US-09-344-510B-1	Sequence 1, Appl
42	708.5	26.8	437	3 US-08-688-988-8	Sequence 8, Appl
43	702.5	26.6	478	4 US-09-107-532A-4922	Sequence 4922, Ap
44	699	26.4	468	4 US-09-489-039A-9557	Sequence 9557, Ap
45	690.5	26.1	485	4 US-09-134-000C-6295	Sequence 6295, Ap

ALIGNMENTS

RESULT 1									
US-08-688-988-37									
; Sequence 37, Application US/0868988B									
; Patent No. 6096545									
; GENERAL INFORMATION:									
; APPLICANT: Lefebvre, Daniel D.									
; APPLICANT: Malboobi, Mohammad A.									
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS									
; FILE REFERENCE: PPL96-03									
; CURRENT APPLICATION NUMBER: US/08/688,988B									
; CURRENT FILING DATE: 1996-07-31									
; NUMBER OF SEQ ID NOS: 48									
; SOFTWARE: FASTSEQ for Windows Version 3.0									
; SEQ ID NO 37									
; LENGTH: 551									
; TYPE: PRT									
; ORGANISM: Prunus serotina									
; US-08-688-988-37									
Query Match									
Best Local Similarity 45.3%; Pred. No. 5.3e-97;									
Matches 213; Conservative 70; Mismatches 171; Indels 16; Gaps 11;									
QY	6	ALPNDEFWGFATAYQIEGAVKEGGRGPSIMDTYCHLEPSR-TNGANGDVACDHYRYDE	64						
DB	42	ALBPGLFPTGTAASAAQFEGAKEDGRGSIWDTYTHNSERIKDGSNGDVAVDQYHRYKE	101						
QY	65	DPDLTKYKGAARFSLWSRIITPLGRLDPVNEEGIEFYSKLDLILRRGITPWTLYH	124						
DB	102	DVIMKMGDAARFSLWSRVLPNGVSGVNEDEGKFTNNILINELRGLKRPVITYH	161						
QY	125	WDLPOLADHRYGWLWVEEVDLDFERYARLCFERFGDVRQVNMITINXPMQAIYVATGS	184						
DB	162	WDLPOLADHRYGWLWVEEVDLDFERYARLCFERFGDVRQVNMITINXPMQAIYVATGS	220						
QY	185	NAPGRSSII-NKSTEGNTATPEWLAGKQAIMSHARAVAVSDPFRSQKQIGISLNGD	242						
DB	221	HAGRCASAMOKLCTGNSATPEPLVTHQDLAAAVKLYKDEYQASQNGLIGITLVP	280						
QY	243	YVPEWMSNERDDEARERMEFHIGFANPIFLAKYVPEEMKQGLERPLPADRAIL	302						
DB	281	WEPD-ASEAEDINAAFRSLDIFFGWPMDBL-TNGVPHLMRSIVGRRLNFTBOSKLL	338						
QY	303	NAGETDFYGNVYTSQFARHLDPVETDYLGAIHHEQNK---DGPVGEESGLAMLR	358						
DB	339	K-GSPDGIQNTYTRVYS--NAPKITSVNASITTPQVATALGKVPICGMAASGLV	395						
QY	359	SCDPMFRKHLARVYGLYKGP-IYITNGCPCEPEENNTCEAVNDPFRIRYFPHSLDIS	417						
DB	396	VYKGIHDLVLYTKKYNDELITYITENGVDPEFNDPKLSMEALKDTRIDFYVRLHCYLO	455						

Best Local Similarity 43.5%; Pred. No. 4,1e-95;
Matches 204; Conservative 76; Mismatches 168; Indels 19; Gaps 12;

QY 6 ALNDNEMGATATAYQIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGDVACHYHYRDE 64
DB 9 SFRGRGTFGASAAAYVEGAMNEGGRGPSIMDTFTDHPDKIADHSNGDVATSYKKYKX 68
QY 65 DFDLLTKYGAARFSLMSRIIPLGRLDPVNEEGIEFYSKLIDALLRRGRTTPTVLYH 124
DB 69 DVTLLDGLDSTYRFSISMSRIIPKGTLOGGINOEGIYNDLINELKRGIRPMITLFI 128
QY 125 WDLPOALHRRYGMVLNVEEVDLPERYARLCEFERGDRVQNMWITTXPMIOAYGATGS 184
DB 129 MDVPOALDEDSYKGRSSEIYN-DFKDYADICFEKFGDRVGMWITLNEPMSLSTMGYAFGR 187
QY 185 NAGRSISINHSIEGNTATPEMLAGKAOIMSHAAVAVYSDRPPSGKQIGISLNGDY 244
DB 188 HARGRSYWGCPAGSANEPEVTHNLHLAHANAAYIRDNKATONGEIGITLNSLMY 247
QY 245 EPWDSNEPRDKEAERMEFHIGMFANPIFLKKDYPSMKOLGERLPALTTPADPAIINA 304
DB 248 EPI-SHSHEVEALTRALDPMFGWMDPL-VNGDYPIPMALVDRLPFTTHAESELIK- 304
QY 305 GETDFYGMNYYTSQFARHLDPV----PRTDYGAIHENKDKGSPVEESGLAMLR 359
DB 305 GSYDFIGINYYTSNVAQH--APVTEDHTPDNSYFDSVNGSGENGVPIGLQG-SWIFY 361
QY 360 CPMFRKHLARVGLYKRP-IYTTENGCPGEGEENMTCEAVNDPFRIRFDSHLSISK 418
DB 362 YPRLGKELLVYRRRCNPFIYITENG--TAEVEKEKGVPLHDPKEXELTYHLAQLV 418
QY 419 AIPODGVVVGYPAMLLDLNLEMSDGYGPRFVTFDY-TTLKRTPKKS 466
DB 419 AI-REGVRYKGFHTMALTDNFPENDKGTTERFGLIYIDYDKOPRQPKDS 466

RESULT 5

US-08-688-988-39
Sequence 39, Application US/08688988B
Patent No. 6096545
GENERAL INFORMATION:
APPLICANT: Lefebvre, Daniel D.
APPLICANT: Malboobi, Mohammad A.
TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
FILE REFERENCE: PRL96-03
CURRENT FILING DATE: 1996-07-31
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 39
LENGTH: 531
TYPE: PRT
ORGANISM: Prunus avium
US-08-688-988-39

Query Match 35.0%; Score 924.5; DB 3; Length 531;
Best Local Similarity 41.1%; Pred. No. 1e-90;
Matches 197; Conservative 86; Mismatches 182; Indels 15; Gaps 11;

QY 11 FEWGFATAAYQIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGDVACDHYARDEDFDLL 69
DB 37 FTGTGTASTAQLEGANIDGRGSPIMDAFTNNHPEKTTDGSNGDVAMIDYHRRKEDVAIM 96
QY 70 TKYGAKAYRSLMSRIIPLGRLDPVNEEGIEFYSKLIDALLRRGRTTPTVLYHMDLQ 129
DB 97 KDMGLAYRFSISMSRIIPKGTLOGGINOEGIYNDLINELKRGIRPMITLFIHMDVPO 156
QY 130 ALMDRYGMLNVEEVDLPERYARLCEFERGDRVQNMWITTXPMIOAYGATGSNAAGR 189
DB 157 ALDEEYGGVLS-PRIVDFKAYIELCYKERGDRVGMWITLNEPFTTISNHGTTGIAHAGR 215
QY 190 SS--INHGTEGNTATEPMLAKGAQIMSHAAVAVYSDRPPSGKQIGISLNGDYEPW 247

DB 216 CSSWYDPTCLGDSGTPEYLVTHNLHLAHAAVKLYREKYAQSQCIEGIVTWSHFEF- 274

QY 248 DSNPEPRDKEAERMEFHIGMFANPIFLKKDYPSMKOLGERLPALTTPADPAIINAGET 307
DB 275 ASBSQCKINASVADALPMFGWMDPL-TRGDDPQSRSLYKERLPFTPEQSSSL-IGSY 332
QY 308 DFGMNYVTSQFA---RHLDGPVPEPTYLGAIHENKDKGSPVEESGLAMLRSCPM 363
DB 333 DYGIVVYGARVASAPBEDYSIPTPSYLTDAVYVWVTELTNGVPIGPOAASDMLVYYPKG 392
QY 364 FRKHLARVGLYKRP-IYTTENGCPGEGEENMTCEAVNDPFRIRFDSHLSISALITQ 422
DB 393 LYDLVLYTKKXNDPIMYITTEGMDENPNPKISLEQALDMSNRIDCYRHLCYLOEAI1- 451
QY 423 DGWVVGYPAMLLDLNLEMSDGYGPRFVTFDYTT-LKRTPKSALVTKDMPAAROV 480
DB 452 EGANVGIVFAMSLDLNLEMSDGYGPRFVTFDYTT-LKRTPKSALVTKDMPAAROV 509

RESULT 6

US-09-306-593-2
Sequence 2, Application US/09306593
Patent No. 6184018
GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
APPLICANT: Chen, Huizhong
APPLICANT: Ximenes, Eduardo A.
TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from
FILE REFERENCE: 31-98us
CURRENT APPLICATION NUMBER: US/09/306,593
CURRENT FILING DATE: 1999-05-06
EARLIER APPLICATION NUMBER: US 60/084,494
EARLIER FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 657
TYPE: PRT
ORGANISM: Orpiniomyces sp. PC-2
US-09-306-593-2

Query Match 34.9%; Score 923.5; DB 3; Length 657;
Best Local Similarity 36.0%; Pred. No. 1.9e-90;
Matches 209; Conservative 82; Mismatches 158; Indels 131; Gaps 17;

QY 7 LPNDFEWFATATAYQIEGAVKEGGRPSIMDTYCHLEP-----SRTNGANGDVACDH 58
DB 74 LPADFCKMGATATAYQIEGAMDEEGRESVMDHPTLHPKXVESGDRSKDFSTGNTACDS 133
QY 59 YHRDEDFDLTKYGAARFSLMSRIIPLG-----GRDPVNEEGIEFYSKLIDALL 112
DB 134 YHRDEDFDLTKYGAARFSLMSRIIPLG-----GRDPVNEEGIEFYSKLIDALL 112
QY 113 RRGITPMTLYHMDLPOALHRRYGMVLNVEEVDLPERYARLCEFERGDRVQNMWITTXP 172
DB 193 KNDIVFVTLVHMDLPOALHRRYGMVLNVEEVDLPERYARLCEFERGDRVQNMWITTXP 251
QY 173 MIOAIYVATG-----SNAP-----GRSSINHSIEGNTATEPMLAKGA 211
DB 252 WNVCVSGYHLPKPKAYRCTGEAPRLQNSTDLDEGGCSYEIGPYOSKNSPFLPANNV 311
QY 212 Q-----IMSHAAVAVYSDRPPSGKQIGISLNGDYEPWMSNEPRDE--- 256
DB 312 POKLEDFWCSHNLILGHAAVYKVRREKFKQKGLIGITVDGAQIPWEPGMTKEYEN 371
QY 257 ---AERMEFHIGMFANPIFLKKDYPSMKOLGERLPALTTPADPAIINAGETDFYGMN 313
DB 372 NLKYANLAAEFKRGWSDPPMV-GDYPKSKYKEMGKDLPEFTBEKILKIGSSSDPLGN 430
QY 314 YVTSQF---ARHLDPVPEPTYLGAIHENKDKG-----PVG 349

Db 431 TYTAHMAQAQKNEDG-----SYIOPTAEANFNSKKOMMDNCKGRGDWTCIPTLIG 485
QY 350 BESSGLAW-----LRSCPDMRKHLARVYGLYGRPIYITENGCPCG----- 390
Db 486 SOAGSSMTKAPITIRVGLNMFSK--RYEGLIKNGIYITENGCQOPYKVARANDEVTK 542
QY 391 -----BENMTCEE-----AVNDPRIRYFSDHLSISKAITODGVVYK 429
Db 543 KYFESIGCPKADYKEDIREDIEDLLEGTLMHDYRIMYOYLKLNLAIAVANDIDVRG 602
QY 430 YFAALLNLNLEMSDGYGRFGVTFYD--TTLKRTPKKS 466
Db 603 YMAWLLDNFEMWEXETRFQWTYIDFYNDKEMKRVKPS 642

RESULT 7

US-09-122-230-7
; Sequence 7, Application US/09122230A
; Patent No. 5973228
; GENERAL INFORMATION:
; APPLICANT: Carlson, et al.
; TITLE OF INVENTION: Coniferin Beta Glucosidase cDNA for Modifying Lignin
; TITLE OF INVENTION: Composition in Plants
; FILE REFERENCE: 50532
; CURRENT APPLICATION NUMBER: US/09/122,230A
; CURRENT FILING DATE: 1998-07-23
; EARLIER APPLICATION NUMBER: U.S. 60/053,566
; EARLIER FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Pinus contorta
US-09-122-230-7

Query Match 34.4%; Score 910; DB 2; Length 513;
Best Local Similarity 41.6%; Pred. No. 3,66-89;

Matches 198; Conservative 83; Mismatches 165; Indels 30; Gaps 13;

QY 8 PNDPEWGATAAVOIGAVKEGGRGPSIMDYCYHLEPSRTGANGDVA CDHYHYDDEDF 67
Db 31 PSDMFEGTASAYOEAGVNRDGGPSTWDLTTHMPGRINDSSNGDVAVDYHRHYMEDIE 90
QY 68 LITTKYAKAVRSLSMSRIIPLGRLDPVNEEGIEFYSKLIDALLRGITPWTLYHMDL 127
Db 91 LMASLGLDAYRFSISWSRIIPLEG--RGEINMAGIEYNNLIDALONGCIOQFVTLFHFDL 148
QY 128 PQALHDRVYGMVNEVQVLDPERYARLCFERFGDRVQNMWITINXPMIOAIYGYATGSNAP 187
Db 149 PKALEDSYGMWLSPOIIN-DEEAYAEICFRAFGRVXKMTAVNEPNLFVPLGYGVGFPP 207
QY 188 GRSS---INKSTSGN-TATEPMLAGKAOIMSHARAVAVYSRDRPSOKOIGISLNGDY 243
Db 208 TRCAAPHANPLCMQNGCSAEPYLAHNVILAHASAEKREKXOKIQGSGISGLVISA PW 267
QY 244 YEPWDSNEPRDKEAEERMEFHIGWFPANPIFLKXDYPSMKKOLGERLPAITPADFAILN 303
Db 268 YEPLLE-NSPERSAVDRILSENLRLWFLDPYF--GDYQEKMERLGRSLPSIS-SELAKL 324
QY 304 AGEDDFYGMNYYTSQFARHLDGVPPE-TDYL-----GAIHEHENDKDSVGESESG 353
Db 325 RGSFDTYMGINYYTTLVYASTPPLSPDHTQYLYPDSRVLYLGERG-----GVSIGERTG 377
QY 354 LAWLRSQCPDMERKHLARVYGLYGRP-IYITENGCPCGEENMTCEAVNDPFRIRYFDSH 412
Db 378 MDGLFVVPHGIQKIVVEYKEFYDNPTIILANGPBESESSSTIQENLNDVRRIRFRFGDC 437
QY 413 LDSISKAITODGVVYKGFAMALIDNLEMSDGYGRFGVTFYD--TTLKRTPKKSA 467
Db 438 LSYLSAAL-KNGSDVRGTYFWVSLDNEFWMAFGYTIRGLVHVPDISQKRYPKLSA 492

RESULT 8

US-09-431-470-4
; Sequence 4, Application US/09431470
; Patent No. 6433249
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Simmons, Carl R.
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Folkerts, Otto
; TITLE OF INVENTION: The Use of Beta-Glucosidase to Enhance
; TITLE OF INVENTION: Disease Resistance to Insects in Crop Plants.
; FILE REFERENCE: 5718-43
; CURRENT APPLICATION NUMBER: US/09/431,470
; CURRENT FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: 60/107,920
; PRIOR FILING DATE: 1998-11-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Zea mays
US-09-431-470-4

Query Match 34.1%; Score 902; DB 4; Length 563;
Best Local Similarity 39.5%; Pred. No. 3,1e-88;
Matches 195; Conservative 98; Mismatches 169; Indels 32; Gaps 15;

QY 8 PNDPEWGATAAVOIGAVKEGGRGPSIMDYCYHLEPSR-TNGANGDVA CDHYHYDDEDF 66
Db 76 PSDIFGAATSAYOIGANMEDKGESNMDFCHNPERIMDSNADIGANSYMYKTIV 135
QY 67 DLTTKYAKAVRSLSMSRIIPLGRLDPVNEEGIEFYSKLIDALLRGITPWTLYHMD 126
Db 136 RLTKEMGMADYRSLSPRLPKYVGGINDGIDYKRLNLNLENGIBEPVTLFHM 195
QY 127 LPQALHDRVYGMVNEVQVLDPERYARLCFERFGDRVQNMWITINXPMIOAIYGYATGS 184
Db 196 VPQALEKKGGLFDKTKQKILVNDYKMFACVCFDNFDKXKNMLTFNEPQTFPSFSYGTGV 255
QY 185 NAGRSS--INKSTSGNATEPMLAGKAOIMSHARAVAVYSRDRPSOKOIGISLNGD 242
Db 256 FARGRSPGIDCAIPTGNSLVBEFYIAGHNILAHAAVLDLYNNKYK-GERGRIGLAFDW 314
QY 243 YEPWDSNEPRDKEAEERMEFHIGWFPANPIFLKXDYPSMKKOLGERLPAITPADFAIL 302
Db 315 GRVPYGTSP-LBQAKERSMDINLGFLEPV--VRGYPFSMSLSARELPFFSDKQOEXL 372
QY 303 NAGETDFYGMNYYTSQFARHLD-----GPVETDYLGAITHEQENKDSVGEESGLAWL 357
Db 373 -VGSYNNMLGINYYTISFSKHIDISPKYSPVLNTDAYSQE-TYGPDKRIGPPMGNPMI 430
QY 358 RSCPDMERKHLARVYGLYGRP-PYITENGCPCG-----EENMTCEAVNDPFRIRYFDS 411
Db 431 YLYPEGLKQIILIMKKNYNPNPYITENGI---GDVDTKEKELPMEEALNDYKRLDYTOR 487
QY 412 HDSISKAITODGVVYKGFAMALIDNLEMSDGYGRFGVTFYD--TTLKRTPKKSA 467
Db 488 HISTLKESIDL-GANHGAFMSLNDNFEMVAGIYERYGIYVVDKRNMT---RYMKESA 543
QY 468 LVLKDMFAARQRYK 481
Db 544 KWLKEPNTAKKPSK 557

RESULT 9

US-08-688-988-38
; Sequence 38, Application US/08688988B
; Patent No. 6096545
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Daniel D.
; APPLICANT: Malboochi, Mohammad A.
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS

FILE REFERENCE: PPL96-03
CURRENT APPLICATION NUMBER: US/08/688,988B
CURRENT FILING DATE: 1996-07-31
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 38
LENGTH: 531
TYPE: PRT
ORGANISM: Manihot esculenta
US-08-688-988-38

Query Match 33.9%; Score 896; DB 3; Length 531;
Best Local Similarity 40.3%; Pred. No. 1,2e-87;
Matches 190; Conservative 90; Mismatches 172; Indels 20; Gaps 11;

QY 8 PNDFFMGFATAAIOIEGAVKEGGRPSIMPTVCHLEPSR-TNGANGVACDHYHARDEDF 66
DB 35 PDDFIFGATSAIOIEGATAGKAGAPVWDIFSKETPDRLDSDGVAADFNRYYIQDI 94
QY 67 DLTLYGAKAYRSLWSRIIPGGRLDPVNNEGIEFYSKIDALLRGITPPWVTLVHMD 126
DB 95 KAVKMGFNAPFRSISMSRVIIPSGRRREGVNEBGIOFVNDVINEIISNGLEPVTIIFHMD 154
QY 127 LPOALHVRGGMVNEBVQIDFERVYARLCFERFGDRVQNMITINXPMIQAIFYATGSNA 186
DB 155 TPQALODKYGGFLS-RDIYVDYQYADLLFERFGDRVYKPMWTFNBPAYVGFAMHDGVFA 213
QY 187 RGGSS--INKHSTEGTATEPMLAGKAQIMSHARAVAVSRDPSPQKQIGISLNDY 244
DB 214 PGCSVWNNQCLAGSATEPYIVAHNLISHAAAVQYKYYOGTQKIGITLFTFWY 273
QY 245 EPDSDNEPRDKAEARMERMEFHIGFANPIFLKKDYPSMKQIGERLTPADFAILNA 304
DB 274 EPL-SOSKYDVQAKTALDPMFGLMCDPMYGR-YRPTWVDLAGDKLIGTDESQULR- 330
QY 305 GETDFYGMNYYISQFAHLDGVPEPTD---YL-GAIEHOENKDGSPVGEESGLAWL 357
DB 331 GSYDFVGLQYTYAYVAE---PIPVDPKPRRYKTDGCVNATPYDLNGNLIGFOAYSSWF 386
QY 358 RSGPDMFRKLAVYGLYCKP-IYITENGCPGCEENMCEAVNDFRIRYDSHLDST 416
DB 387 YIFPKGIRHFLNLTQYNDPVLYVTEGNDVNNESQPIEALQDDPFIYSTYKKHMMNA 446
QY 417 SKAITODGVVVKGFAMALDNLNLEWSDGYGPRFGVTFTDY-TLLKRTPKKSA 467
DB 447 LGSLLKMYGVKLCGYFAMSYLDNFMENIGYISRGGLYVYKNNLTRYPKKSA 498

RESULT 10
US-08-914-375C-72
Sequence 72, Application US/08914375C
Patent No. 6377893
GENERAL INFORMATION:
APPLICANT: Steven A. Benner
Applications of Protein Structure Predictions
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven A. Benner
STREET: 1501 NW 68th Terrace
CITY: Gainesville
STATE: FL
COUNTRY: United States
ZIP: 32605-4147
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/914,375C
FILING DATE: 19-Aug-1997
CLASSIFICATION: 702/20
TELECOMMUNICATION INFORMATION:

TELEPHONE: 352 392 7773
TELEFAX: 352 331 0462
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 459
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: amino acid
ORIGINAL SOURCE:
ORGANISM: Trifolium repens
FEATURE:
OTHER INFORMATION: bgl's trip non-cyanogenic strand-glucosidase precursor
SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-08-914-375C-72

Query Match 33.3%; Score 880.5; DB 4; Length 459;
Best Local Similarity 40.3%; Pred. No. 4,6e-86;
Matches 186; Conservative 80; Mismatches 177; Indels 19; Gaps 8;

QY 6 ALPNDDEMGFATAAIOIEGAVKEGGRPSIMPTVCHLEPSR-TNGANGVACDHYHARDE 64
DB 5 SFRPGTIFGAGSAYVPEGAVNEGGRPSIMPTVTHKYEKIRDSNADITVDQHYRYKE 64
QY 65 DFDLTYGAKAYRSLWSRIIPGGRLDPVNNEGIEFYSKIDALLRGITPPWVTLVH 124
DB 65 DVGIMKDQNMDSYRFSISMPRIPLKGLSGINHEGIKYNNMLINELLANGIQPFVTLF 124
QY 125 WDLPOLHVRGGMVNEBVQIDFERVYARLCFERFGDRVQNMITINXPMIQAIFYATGS 184
DB 125 WDLPOLHVRGGMVNEBVQIDFERVYARLCFERFGDRVQNMITINXPMIQAIFYATGS 183
QY 185 NAFGRSSINKHSTEGTATEPMLAGKAQIMSHARAVAVSRDPSPQKQIGISLNDY 244
DB 184 NAFGRSSINKHSTEGTATEPMLAGKAQIMSHARAVAVSRDPSPQKQIGISLNDY 243
QY 245 EPDSDNEPRDKAEARMERMEFHIGFANPIFLKKDYPSMKQIGERLTPADFAILNA 304
DB 244 MPDSDNSIDIKAKAESLDFQGLFMEQL-TTGDSYKSNRRIYQNLPRFSKRESSLVN- 301
QY 305 GETDFYGMNYYISQFAHLDG-----PVPEPTDVLGAIEHOENKDGSPVGEESGLAWL 359
DB 302 GSYDFYGINVYSSYSNAPSHGNAPKSYSTPMINI---SEFKGIPGGRPAASTIWIY 358
QY 360 CPDMFRKLAVYGLYCKP-IYITENGCPGCEENMCEAVNDFRIRYDSHLDST 413
DB 359 YIFPKGIRHFLNLTQYNDPVLYVTEGNDVNNESQPIEALQDDPFIYSTYKKHMMNA 418
QY 414 DSISKAITODGVVVKGFAMALDNLNLEWSDGYGPRFGVTFTD 455
DB 419 YYIRSAI-RAGSNVKGFYAMSLDCNEMWAFGTVRGLNFVD 459

RESULT 11
US-09-431-470-2
Sequence 2, Application US/09431470
Patent No. 6433249
GENERAL INFORMATION:
APPLICANT: Duvick, Jon
Applications of Protein Structure Predictions
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Craeta, Oswald R.
STREET: Folkerts, Otto
CITY: Gainesville
STATE: FL
COUNTRY: United States
ZIP: 32605-4147
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/431,470
FILING DATE: 1999-11-01
PRIOR APPLICATION NUMBER: 60/107,920
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 566
TYPE: PRT
TELECOMMUNICATION INFORMATION:

ORGANISM: Zea mays
US-09-431-470-2

Query Match 33.2%; Score 879; DB 4; Length 566;
Best Local Similarity 38.3%; Pred. No. 9,4e-86;
Matches 192; Conservative 94; Mismatches 183; Indels 32; Gaps 14;

1 MPESIALPNDFEWGFPATAYOIGAVKEGGRPSIMDTYCHLEPSR--TGANGDVACDY 59
2 IPRDMFSDFTFGAATSAVOIGAWNEDGSGSNMDFCHNPERLILDSNDIDANFY 131
3 HRYDEDFLLTKYGAAYRFSLSWSRIIPLGRIIDPVNEBIEFYSKULIDALIRGITPW 119
4 HMYKTDLVRLKEMGMDAYRFSISMPRIIPKGTKEGGINPOGIKYRMLIMLLENGIEPY 191
5 VLYHMDLPQALHRYGGMVNEEVOI--DPERYARLCFERFGDGVONMWTINXPMIOAI 177
6 VTIFFMDVPOALEEKYGGFLDKSHKSIYEDYTYPAKCEDFGDKVKNMILTENEPQTFIS 251
7 YGIATGSAAPGSS--INKHSTEGNTATEPMLAGKAOIMSHARAVALVAVSRDFRPSQKQI 235
8 FSYGTGVFAPGKCSGLDCAIPTGSLVPEYTAGINILLAAEAVDLYNKHKKDDT-R1 310
9 GSIINGDYIEPWSNEPRDKEAERMEFHIGWFPANPIFLKQDYPESKQOLGERLPALT 295
10 GLAFDVGWVPGTGF-LDKQAEERSWDINLGWLEPV-VRGDYPFSMRSLARELRLEPFK 368
11 PADFAIANGERTDFYGMNYYTTSOPARHLD-----GPVETDYIGAHEHOKKSGPVGE 350
12 DEQEKEL-AGSYNMGILNYYTSRFSKNIDISPNYSPVLTNTDAVNSQE-VNPDQKPIGP 426
13 ESGLAWLRSCPDWFRKHLARVYGLYGR-PITYTENG-----CPCGEENMTCEAVN 401
14 PKMNPITVYVPELKKOLIMMKKYGNPPIYITENGIGDVTKEPPLP-----MEALIN 480
15 DPFRIIRFSDHLSISKAITQGVVYKGFAMALLDNLMSDGYGRFSGVTFTD-YTLIK 460
16 DYKRLDYIQRHITYLTLESIDL-GSNVQGFASLNDNFEMFAGFTERYGIYVVDNNNCT 539
17 RTPKKSALVLDKMFARQRYK 481
18 RYKESAKMLKEPNTAKKPSK 560

RESULT 12
US-08-688-988-34
Sequence 34, Application US/08688988B
Patent No. 6096545
GENERAL INFORMATION:
APPLICANT: Lefebvre, Daniel D.
TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
FILE REFERENCE: PPL96-03
CURRENT APPLICATION NUMBER: US/08/688, 988B
CURRENT FILING DATE: 1996-07-31
NUMBER OF SEQ ID NOS: 48
SOFTWARE: fastseq for Windows Version 3.0
SEQ ID NO 34
LENGTH: 524
TYPE: PRT
ORGANISM: Arabidopsis Thaliana
US-08-688-988-34

Query Match 32.7%; Score 863.5; DB 3; Length 524;
Best Local Similarity 39.8%; Pred. No. 3.9e-84;
Matches 196; Conservative 73; Mismatches 198; Indels 25; Gaps 13;

6 ALPNDEMGFATAAYOIGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHRYDED 65
7 SPEEGELFGTATTAAYVEGAINETCRGPAIMDIYCRARYPERCNDNDGVAVDFHRYED 98
8 FDLITRYGAKAYRFSISMSRIIPLGRIIDPVNEBIEFYSKULIDALIRGITPMWTLVHW 125

99 IOLMKALNTDAPFMSIAMPRIIPPHGRKEKGVSOAGVQFYHDLIDELIKNGITPFWTFHH 158
100 DLPQALHRYGGMVNEEVOIDFERYARLCFERFGDGVONMWTINXPMIOAIYVATGSN 185
101 DTPQDLDEYVGGLFS-ERIVKDFREYADVFQYGGKVKHMITNEPFWFLHAGYVGKK 217
102 ARGSS--INKHSTEGNTATEPMLAGKAOIMSHARAVALVAVSRDFRPSQKQIGISANGY 243
103 APGRCSYVNAKKODGSGYEAVLYTHNLLISHAEVAV-RKCEKCKGKIGIAHSPAM 276
104 YEPWDSNEPRDKEAERMEFHIGWFPANPIFLKQDYPESKQOLGERLPALTADPAIIN 303
105 FEADHLADSOOGASIDALDFILGMHLDITTF-GDYPQIKMDIYGHRLPFKFTTEQKAKK 335
106 AGEFDYGMNYYTTSOPARHLDGPVE-----TDYIGAHEHOKKSGPVGEESGLAMLR 358
107 A-STDFVLNYYTTSVFSNHLKPPDSKPRMWOBSL-ITMESKNAQVAYAGSPFLTALN 392
108 SCPDWRKHLARVYGLYGR-PITYTENGCPCEGEENMTCEAV-----NDPFRIIRYFDSHL 413
109 VYSRGFRSLKXIKDKYANPEIMENGY--GEE-LGASDSVAVGTADHNRKTYLQRL 448
110 DISKAITQDGVVYKGFAMALLDNLMSDGYGRFSGVTFTDY-TTLKRTPKKSALVLDK 472
111 LMOEAVCIDKVNVTGYFVWSLNDNFEMQDKRFRGLYYVDFKONILTRYEKESGKYKD 508
112 MFAARQRYKVA 484
113 FLS--QVAPSA 518

RESULT 13
US-09-914-841A-1
Sequence 1, Application US/09914841A
Patent No. 6645750
GENERAL INFORMATION:
APPLICANT: Amaro Enzyme Inc.
TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE
FILE REFERENCE: 066072
CURRENT APPLICATION NUMBER: US/09/914, 841A
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: P. Hel. 11-056299
PRIOR FILING DATE: 1999-03-04
PRIOR APPLICATION NUMBER: PCT/JP00/01242
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 429
TYPE: PRT
ORGANISM: Camelilia var. sinensis
US-09-914-841A-1

Query Match 32.6%; Score 861; DB 4; Length 429;
Best Local Similarity 42.6%; Pred. No. 5.2e-84;
Matches 184; Conservative 73; Mismatches 159; Indels 16; Gaps 11;

47 TNGANGDVACDHYHYRDEDFLLTKYGAAYRFSLSWSRIIPLGRIIDPVNEBIEFYSK 106
48 SNGSTGVADDFHYRYKEDVAVLKFIIGLDFRMSISARVLPKGLSGVAKKEGIARFYN 61
49 LIDALLRGTITPWWTLVHMDLPQALHRYGGMVNEEVOIDFERYARLCFERFGDGVONM 166
50 VINDLISKGIQFIRITFHMDDLPOALEDEYGGFLSPHVN-DFRDPAELCFKEKEDRKYHW 120
51 ITNNEPWSYSYGGYDAGLAPKCSAFMAFCPKNSGTEPIYVTHNLLISHAAVVKYKE 180
52 ITINXPMIOAIYVATGSNAPGR-SSINKHSTEGNTATEPMLAGKAOIMSHARAVALVAVSR 225
53 DFRPSQKQIGISLNGDYIEPWSNEPRDKEAERMEFHIGWFPANPIFLKQDYPESKQK 285
54 KYAYQKQIGITIVTYWMIYF-SNSKADKDAACRALDFMGWFIPLSF-GEYPKSMKR 238

Db	400	GNQYYPKGIYYM--DYFKTK-----YGDPLIYVTENGFSFPSSENR--EQAIADY	447
QY	404	FRIRYPSHSDSISKAITODGVVKGYPFAMALLDNLEMSDGYGPRFGVFPTDYTL-KRT	462
Db	448	KRIDYLCSHLCFLRKVIKEKGVNVRGTFAMALGDNYIEFCCKGFTVRFGLSYVNWEDLDNRN	507
QY	463	PKKS	466
Db	508	LKES	511

Search completed: March 25, 2004, 15:57:43
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2004, 15:56:42 ; Search time 46 Seconds
(without alignments)
2753.137 Million cell updates/sec

Title: US-10-026-140-2
Perfect score: 2644
Sequence: 1 MPESLALPNDPFWGFAATAY.....KSLVLKDMFAARQYKVA 484

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 26161801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppa/PC7_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppa/PC7US_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2642	99.9	484	US-10-026-140-2	Sequence 2, Appl1
2	1210	45.8	490	US-10-369-499-13405	Sequence 13405, A
3	1058.5	40.0	491	US-10-424-599-169310	Sequence 169310, A
4	1058.5	40.0	511	US-10-425-114-43168	Sequence 43168, A
5	999.5	37.8	515	US-10-425-114-55300	Sequence 55300, A
6	999.5	37.8	516	US-10-424-599-243532	Sequence 243532, A
7	999.5	37.8	519	US-10-425-114-48403	Sequence 48403, A
8	999.5	37.8	519	US-10-425-114-55444	Sequence 55444, A
9	999.5	37.8	527	US-10-425-114-56976	Sequence 56976, A
10	999.5	37.8	527	US-10-425-114-70165	Sequence 70165, A
11	999.5	37.8	527	US-10-425-114-71098	Sequence 71098, A
12	997	37.7	523	US-10-424-599-213409	Sequence 213409, A
13	997	37.7	552	US-10-425-114-55483	Sequence 55483, A
14	996.5	37.7	503	US-10-424-599-147332	Sequence 147332, A
15	986	37.3	510	US-10-424-599-219558	Sequence 219558, A

16	986	37.3	522	US-10-425-114-48576	Sequence 48576, A
17	977	37.0	531	US-10-424-599-214089	Sequence 214089, A
18	972	36.8	525	US-10-425-114-50107	Sequence 50107, A
19	963	36.4	521	US-10-425-114-59902	Sequence 59902, A
20	961.5	36.3	515	US-10-424-599-175202	Sequence 175202, A
21	960	36.3	528	US-10-425-114-65044	Sequence 65044, A
22	959	36.3	531	US-10-425-114-40755	Sequence 40755, A
23	956.5	36.2	526	US-10-424-599-213328	Sequence 213328, A
24	953	36.0	524	US-10-424-599-281853	Sequence 281853, A
25	943.5	35.7	545	US-10-195-781A-4	Sequence 4, Appl1
26	941.5	35.6	506	US-10-424-599-181025	Sequence 181025, A
27	934	35.3	584	US-10-425-114-64343	Sequence 64343, A
28	930.5	35.2	574	US-10-425-114-44487	Sequence 44487, A
29	928.5	35.1	527	US-10-424-599-283766	Sequence 283766, A
30	925	35.0	523	US-10-425-114-54197	Sequence 54197, A
31	922	34.9	541	US-10-425-114-50689	Sequence 50689, A
32	914	34.6	510	US-10-425-114-45687	Sequence 45687, A
33	914	34.6	512	US-10-425-114-44614	Sequence 44614, A
34	914	34.6	513	US-10-424-599-263085	Sequence 263085, A
35	914	34.6	513	US-10-425-114-45855	Sequence 45855, A
36	912.5	34.5	565	US-10-425-114-38182	Sequence 38182, A
37	902	34.1	478	US-10-425-114-54717	Sequence 54717, A
38	884	33.4	508	US-10-425-114-46495	Sequence 46495, A
39	884	33.4	567	US-10-425-114-65966	Sequence 65966, A
40	884	33.4	571	US-10-425-114-72958	Sequence 72958, A
41	884	33.4	571	US-10-425-114-73036	Sequence 73036, A
42	884	33.4	572	US-10-425-114-64277	Sequence 64277, A
43	884	33.4	572	US-10-425-114-72696	Sequence 72696, A
44	882	33.4	571	US-10-425-114-66650	Sequence 66650, A
45	879	33.2	567	US-10-425-114-65837	Sequence 65837, A

ALIGNMENTS.

RESULT 1
US-10-026-140-2
; Sequence 2, Application US/10026140
; Publication No. US20030114330A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedgebaur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: US/10/026,140
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ. ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Trichoderma reesei
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(484)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-10-026-140-2

Query Match 99.9%; Score 2642; DB 14; Length 484;
Best Local Similarity 100.0%; Pred. No. 9.4e-261;
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPESLALPNDPFWGFAATAYQIEGAVGEGRGPSINDTYCHLEPSRTNGANGVACDHYH 60
Db 1 MPESLALPNDPFWGFAATAYQIEGAVGEGRGPSINDTYCHLEPSRTNGANGVACDHYH 60
QY RYEDDFLLTKYAKAYRFLSLMSRIIPLOGRDPVNEBIEFFYSKIDLLRGRTPWV 120
Db 61 RYEDDFLLTKYAKAYRFLSLMSRIIPLOGRDPVNEBIEFFYSKIDLLRGRTPWV 120
QY RYEDDFLLTKYAKAYRFLSLMSRIIPLOGRDPVNEBIEFFYSKIDLLRGRTPWV 120
Db 61 RYEDDFLLTKYAKAYRFLSLMSRIIPLOGRDPVNEBIEFFYSKIDLLRGRTPWV 120

QY 121 TLVHMDLFOALHRYGMLNVEVQDLEFERYARLCEFERFGDRVQNMITINXPMIOAIYGY 180
Db 121 TLVHMDLFOALHRYGMLNVEVQDLEFERYARLCEFERFGDRVQNMITINXPMIOAIYGY 180
QY 181 ATGSAAPRRSSINKESTGNTATETPMLAGKQIMSHAAVAVYSDFPSPQKQIGISLN 240
Db 181 ATGSAAPRRSSINKESTGNTATETPMLAGKQIMSHAAVAVYSDFPSPQKQIGISLN 240
QY 241 GDYEPWMSNPRDKEAERMEFHIQWFPANPIFLKKQYPSMCKQJGEBRLPALTPADFA 300
Db 241 GDYEPWMSNPRDKEAERMEFHIQWFPANPIFLKKQYPSMCKQJGEBRLPALTPADFA 300
QY 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQENKQDSPVGEESGLAMLRSC 360
Db 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQENKQDSPVGEESGLAMLRSC 360
QY 361 PDMPKHLARVYGLYGRPIYITENCCPCPGSENMTCBAVNDPFRIRFDSHLSISKAI 420
Db 361 PDMPKHLARVYGLYGRPIYITENCCPCPGSENMTCBAVNDPFRIRFDSHLSISKAI 420
QY 421 TODGVVVGYPFAMALLDNLEMSDGYGPRFGVTFTDTTLKRTPKKSALVLDMPAARORV 480
Db 421 TODGVVVGYPFAMALLDNLEMSDGYGPRFGVTFTDTTLKRTPKKSALVLDMPAARORV 480
QY 481 KVAA 484
Db 481 KVAA 484

RESULT 2

US-10-369-493-13405
; Sequence 13405, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13405
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(490)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-13405

Query Match 45.8%; Score 1210; DB 15; Length 490;
Best Local Similarity 47.3%; Pred. No. 2,1e-114;
Matches 230; Conservative 82; Mismatches 154; Indels 20; Gaps 4;

QY 7 LPNDFEWGFATAAYQIEGAVGEGRGPISIMDTYCHLEPSRTNGANGVACDHRYRDEDF 66
Db 7 LPNDFEWGFATAAYQIEGAVGEGRGPISIMDTYCHLEPSRTNGANGVACDHRYRDEDF 66
QY 67 DLTKYGAKAARFSLMSRIIPLGRLDPVNEEGIEFYSKIDALLRRGITPMVTLYHW 126
Db 67 ALMSYGVNAVYRFLSWSRIIPLGADDPVNEOGIKYYQDVLDELNNGITPFTLPHWD 120
QY 127 LPQALHRYGMLNVEVQDLEFERYARLCEFERFGDRVQNMITINXPMIOAIYGYATGSNA 186
Db 127 LPQALHRYGMLNVEVQDLEFERYARLCEFERFGDRVQNMITINXPMIOAIYGYATGSNA 186
QY 121 VPQALHRYGMLNVEVQDLEFERYARLCEFERFGDRVQNMITINXPMIOAIYGYATGSNA 180
Db 121 VPQALHRYGMLNVEVQDLEFERYARLCEFERFGDRVQNMITINXPMIOAIYGYATGSNA 180

QY 187 PGRSSINKHSTEGNTATEPMLAGKQIMSHAAVAVYSDFPSPQKQIGISLNQDYEP 246
Db 187 PGRSSINKHSTEGNTATEPMLAGKQIMSHAAVAVYSDFPSPQKQIGISLNQDYEP 246
QY 247 WDSNEPPDKAERMEFHIQWFPANPIFLKKQYPSMCKQJGEBRLPALTPADFAIINAGE 306
Db 247 WDSNEPPDKAERMEFHIQWFPANPIFLKKQYPSMCKQJGEBRLPALTPADFAIINAGE 306
QY 307 TDFPGMNYYSQFARHLDPVPETDYLGAIHEHQENKQDSPVGEESGLAMLRSCQPMDFRK 366
Db 307 TDFPGMNYYSQFARHLDPVPETDYLGAIHEHQENKQDSPVGEESGLAMLRSCQPMDFRK 366
QY 367 HLAARVGLYGRPIYITENCCPCPGSENMTCBAVNDPFRIRFDSHLSISKAI10DGVV 426
Db 367 HLAARVGLYGRPIYITENCCPCPGSENMTCBAVNDPFRIRFDSHLSISKAI10DGVV 426
QY 427 VKGYFAMALLDN-----LEMSDGYGPRFGVTFTDTTLKRT--PKKSAL 468
Db 427 VKGYFAMALLDN-----LEMSDGYGPRFGVTFTDTTLKRT--PKKSAL 468
QY 469 VLKMF 474
Db 469 VLKMF 474

RESULT 3

US-10-424-599-169310
; Sequence 169310, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihui
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 169310
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123901C.1.pcp
US-10-424-599-169310

Query Match 40.0%; Score 1058.5; DB 12; Length 491;
Best Local Similarity 46.7%; Pred. No. 6.5e-99;
Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

QY 8 PNDPEWGFATAAYQIEGAVGEGRGPISIMDTYCHLEPSRTNGANGVACDHRYRDEDF 67
Db 8 PNDPEWGFATAAYQIEGAVGEGRGPISIMDTYCHLEPSRTNGANGVACDHRYRDEDF 67
QY 25 PNFIFGVATSAQIEBACGEGGRSIPDAFTHTEGKILDKANGVAVNHYRWEDID 84
Db 25 PNFIFGVATSAQIEBACGEGGRSIPDAFTHTEGKILDKANGVAVNHYRWEDID 84
QY 68 LTRKYGAKAARFSLMSRIIP--LGRRLDPVNEEGIEFYSKIDALLRRGITPMVTLYHW 125
Db 68 LTRKYGAKAARFSLMSRIIP--LGRRLDPVNEEGIEFYSKIDALLRRGITPMVTLYHW 125
QY 85 LIAKLGFDAYRFSISKRITPDLGTRK--INDEGITFTYNNITNGLERGIGQYVTLIHW 141
Db 85 LIAKLGFDAYRFSISKRITPDLGTRK--INDEGITFTYNNITNGLERGIGQYVTLIHW 141
QY 126 DLFOALHRYGMLNVEVQDLEFERYARLCEFERFGDRVQNMITINXPMIOAIYGYATGSN 185
Db 126 DLFOALHRYGMLNVEVQDLEFERYARLCEFERFGDRVQNMITINXPMIOAIYGYATGSN 185
QY 142 DLPLHLHESWGMVLRKQILEY--PAVYADTCFASFGDRVQNMITINXPMIOAIYGYATGSN 200
Db 142 DLPLHLHESWGMVLRKQILEY--PAVYADTCFASFGDRVQNMITINXPMIOAIYGYATGSN 200
QY 186 APGRSSINKHSTEGNTATEPMLAGKQIMSHAAVAVYSDFPSPQKQIGISLNQDYEP 245
Db 186 APGRSSINKHSTEGNTATEPMLAGKQIMSHAAVAVYSDFPSPQKQIGISLNQDYEP 245
QY 201 APGRRE-----NSLIEPYLAHHQILAAHAAVAVYSKRYDKKGGVGFVDDDEMAE 252
Db 201 APGRRE-----NSLIEPYLAHHQILAAHAAVAVYSKRYDKKGGVGFVDDDEMAE 252
QY 246 PMSNPRDKEAERMEFHIQWFPANPIFLKKQYPSMCKQJGEBRLPALTPADFAI--LNA 304
Db 246 PMSNPRDKEAERMEFHIQWFPANPIFLKKQYPSMCKQJGEBRLPALTPADFAI--LNA 304
QY 253 A-NSDKIEDSAARLDPOLGFLHPIYV--GDYEPVEMERLQDQPKSEBPKLILNA 310
Db 253 A-NSDKIEDSAARLDPOLGFLHPIYV--GDYEPVEMERLQDQPKSEBPKLILNA 310
QY 305 GETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQ--ENKQDSPVGEESGLAMLRSCP 361
Db 305 GETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQ--ENKQDSPVGEESGLAMLRSCP 361

```

Db      311  --LDPIGLNHTSRFISHVTECAEENHY--KVQEMERIVWEGQALGKASRMLYVVP 367
Qy      362  DMFRKHLARVYGLYGRPIYITTEGCGPCGSENMTCBAVNDPFRIRYFDHLSISAKIT 421
Db      368  WGLRKILINYSOKYATPIFTENGMDDEDNDLPLHMLDKLVRFPKGLASVAQAI- 426
Qy      422  QDGVVVGYPFAMALLDNLWMSDGYGPRFGVTFDYTT-LKRTPKKSA 467
Db      427  KDGADVGVGFAMSLDNLFEWAQGYTKRFGLVYVDYKGLSRHPKSSA 473

```

RESULT 4

```

US-10-425-114-43168
; Sequence 43168, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43168
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700748475_FLI.pep
US-10-425-114-43168

```

Query Match 40.0%; Score 1058.5; DB 12; Length 511;

Best Local Similarity 46.7%; Pred. No. 6.9e-99; Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

```

Qy      8  PNDPEMGFATAAYOIEGAVEGGRPSIMDTYCHLEPSRTNGANGVACDHRYRDEDFD 67
Db      45  PNFIFGVATSAAYOIEGACKEGGRPSIMDAFTHTGKILDKSNGDVAVNHYRHYMEDID 104
Qy      68  LTRKYGAKARFSLWSRIIP--LGRDLDPVNEEGIEFYSKLDALLRIGITPWTLYH 125
Db      105  LIAKLGDAAYRFSISWSRIIPDGLGTR--INDEGITFYNNIINGLERGIQPVTLYHW 161
Qy      126  DLFOALHDRYGMVNEVEQDLDFERYARLCFERFGDRVQNMWITINXPMQAIYGYATGSN 185
Db      162  DPLHLHESGWLNKQIIEY-FAVYADTGFASGADVAKWITINELQIYAVNGDYAIF 220
Qy      186  AGRSSINXKSTEGNTATEPMLAGKAQIMSHARAVALYSDRPSQKQIGISLNGDYE 245
Db      221  AGRRR-----NSLIEPYLAHQILAHAAVSIYRSKYKQKQGGYGVVDDCEMAE 272
Qy      246  PMSNERDKEAERMEFHIGWPNPIFLKDYPSMKQQLGERLPALTPADPAI-LNA 304
Db      273  A-NSDKIEDSAARRLDFOLGWFPLHPLY-GDPEVTRSRRLGQLPKFSBEDKILLNA 330
Qy      305  GETDFYGMNYTSQFARHLDPVETDYLGAIHEHQ--ENKQDSPVGESSGLAWLSCP 361
Db      331  --LDPIGLNHTSRFISHVTECAEENHY--KVQEMERIVWEGQALGKASRMLYVVP 387
Qy      362  DMFRKHLARVYGLYGRPIYITTEGCGPCGSENMTCBAVNDPFRIRYFDHLSISAKIT 421
Db      388  WGLRKILINYSOKYATPIFTENGMDDEDNDLPLHMLDKLVRFPKGLASVAQAI- 446
Qy      422  QDGVVVGYPFAMALLDNLWMSDGYGPRFGVTFDYTT-LKRTPKKSA 467
Db      447  KDGADVGVGFAMSLDNLFEWAQGYTKRFGLVYVDYKGLSRHPKSSA 493

```

RESULT 5

```

US-10-425-114-55300
; Sequence 55300, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55300
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: JC-GMFL02220131H02_FLI.pep
US-10-425-114-55300

```

Query Match 37.8%; Score 999.5; DB 12; Length 515;

Best Local Similarity 44.4%; Pred. No. 7.5e-93; Matches 212; Conservative 74; Mismatches 163; Indels 29; Gaps 11;

```

Qy      6  ALPNDPEMGFATAAYOIEGAVEGGRPSIMDTYCHLEPSR-TNGANGVACDHRYRDE 64
Db      39  SPFAGTFPGASSAYOIEGAKKEYGROPSIMDTFQNHQPKIXDGTNGDRLDQYHYKE 98
Qy      65  DPLLTRKYGAKARFSLWSRIIP--LGRDLDPVNEEGIEFYSKLDALLRIGITPWTLYH 124
Db      99  DVOIKGMNLDAYRFSISWSRIIPNGKLSGGINREGINYNNLIHELOTGKLPVTLFH 158
Qy      125  WDLPOLHDRYGMVNEVEQDLDFERYARLCFERFGDRVQNMWITINXPMQAIYGYATGS 184
Db      159  WDLPOLHENEYKQPL-ESTIDFGDYAKCFEFGDRVGMWTFMEPHIFSSHGYAYGT 217
Qy      185  NABGRSSINXKSTEGNTATEPMLAGKAQIMSHARAVALYSDRPSQKQIGISLNGDYE 244
Db      218  KARGKRSQGLRPSG--GTEPRVSHNILLAHAKAQLVYNSYKESQNGSIGITLDSRMF 275
Qy      245  EPWDSNEPRDKEAERMEFHIGWPNPIFLKDYPSMKQQLGERLPALTPADPAI-LNA 304
Db      276  VPY-SDASSDIEATERALDFELICFMFEPPLSGK-YPESMQLYVGRRLPEFSKEE-DELVR 332
Qy      305  GETDFYGMNYTSQFARHLDPVETDYLGAIHEHQ-----NKQDSPVGE 351
Db      333  GSFDFIGLNYTNTYK-----VATGYTDSVHHHDDLSTDPNVLGLTRLNGSSPFGV 386
Qy      352  SGLAWLSRCPDMFRKHLARVYGLYGRPIYITTEGCGPCGSENMTCBAVNDPFRIRYFD 410
Db      387  PGLGWLGVYPKGIRELLIRKLNLYNNPLIYITTEGNGINELDDPLTSPSESLMDYRIDHY 446
Qy      411  SHLDSISKAITQDGVVVGYPFAMALLDNLWMSDGYGPRFGVTFDY-TTKRTPKKSA 467
Db      447  RHLLNVDYAI-RDGVVVGKGFVMSLDCFEMSNGYIPRFGLIPVDHKNNINRSPLKLSA 503

```

RESULT 6

```

US-10-424-599-243532
; Sequence 243532, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with

```



```

Db      103 DVQIMKMNLDAYRFSISMSRILPNGKLSGGINREGINYNNLIHELQTKGLKRPVTLFH 162
Qy      125 WDLPLQALHDIYGGMLNVEEVLQDFERYARLCFERFGRVONMTITXKPMIQAITYATGS 164
Db      163 WDLPLQALENYKGFSL-ESI1DDGDYAKCFEEFGVRVHMFTFENPHI FSSGVAYGT 221
Qy      185 NAGRSSINHGSTEGRNATEPMLAGKAQIMSHARAVALVSRDPRPSQKQIGISLNDY 244
Db      222 KAGGRKSQGRAPDSG--CTEPYVSHNILLAKAKAVOLYNSYKESNGEIGITLDRMF 279
Qy      245 EPDINSNEPRDKEAERMEFHIGFANPIFLKKDYPSMKKQIGERLPALTPADFAILNA 304
Db      280 VPY-SDASDIEATERALDEIGEFMFELTSGK-YEPESQLYVGRRLPEFSKEE-AELVR 336
Qy      305 GETDFYGMNYYTSQFAHLDGPVETDYLGAIHEHQS-----NKGSPIGEE 351
Db      337 GSEDFIGLNTYTTNTAR-----VATGYTDSVHHHPDLSTDPVELGLTRLNGSSSPGPV 390
Qy      352 SGLAMLRSCPMDFKHLARVYGLYKRP-IYTENGCCPGEENMTCEAVNDPFRIRYFD 410
Db      391 PGLQMLCVYRKGIRELLRLKLNLYNPLIYTENGINELDDPTLSPEBSLMDPFRIDYHY 450
Qy      411 SHLDISSKAITODGVVYKGFANALLDNLEMSDGYGRFVGTFTDY-TTLKRTPKSA 467
Db      451 RHLLNVDTAL-RDGVRYKGYFVMSLDCFEWSNGYIRFGILFVDHKNLNRSPKLSA 507

```

RESULT 9

```

US-10-425-114-56976
; Sequence 56976, Application US/10425114
; Publication No. US20040034888A1
GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ. ID NOS: 2003-04-28
; SEQ ID NO 56976
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17071E02_FLI.pep
US-10-425-114-56976

```

Query Match 37.8%; Score 999; DB 12; Length 527;

Best Local Similarity 42.8%; Pred. No. 8.8e-93;

Matches 208; Conservative 86; Mismatches 148; Indels 44; Gaps 15;

```

Qy      6 ALPNDPEWGATATAYQIEGAVKEGGRGSPSTWDTYCHLEPSTNGANDVACHYHRDED 65
Db      64 SFPKGFVFGTATSAVYQVEGAALSTNGRGPSTWDSFAHVPAGINAGNQNDVAVDQYHRKED 123
Qy      66 FDLITTKGAKAYRFSISMSRIIPLG-GRLDPVNEEGIEFYSKLIIDALLRGTTPWTLVH 124
Db      124 VDLMKSLNPAFYRFSISMSRIIPDGEGK---VNDGVAAYNNLINVLLQGSMTPIYLVH 180
Qy      125 WDLPLQALHDIYGGMLNVEEVLQDFERYARLCFERFGRVONMTITXKPMIQAITYATGS 184
Db      181 YDLPLALEKKYGGMLSAKAMD-L-FTYADFCFTYGRVHKMFTFENPRIVALLGYDTGS 239
Qy      185 NAGRSSINHGSTEGRNATEPMLAGKAQIMSHARAVALVSRDPRPSQKQIGISLNDY 244
Db      240 NPORCT--RCAAGGNSATEPYIVAHNFLAHTAVARVRYTKYQAAKGVIGVLDENWY 297

```

```

Qy      245 EPDINSNEPRDKEAERMEFHIGFANPIFLKKDYPSMKKQIGERLPALTPADFAILNA 304
Db      298 EAL-TNSPDDQAAAGARDFHIGFVDP-L-INGHYQIMQDLVKERLPRT-P-EQAKLVK 354
Qy      305 GETDFYGMNYYTSQF--ARHLDGPV-----ETDYLGAIHEHQSNGKSPGEESSGLA 355
Db      355 GSADYIGINEYTSYKMGQVLQVAPSSYSADWQVQVFA-----RNGKPIGPANSK 407
Qy      356 MLRSCPMDFKHLARVYGL-----YGRP-IYTENGCCPGEENMTCEAVNDPFRIR 407
Db      408 WLTIAP-----TGMGCNNYLKEKYGNTIYTENGNDQPS--NLTRDQYLRDATTR 458
Qy      408 YPDSHLSISKAITODGVVYKGFANALLDNLEMSDGYGRFVGTFTDYTLKRTPKSA 467
Db      459 FYRSYIGQKKALDQ-GANVAGYFAMSLDNFEMLAGYSSKFGIYVVDFTLRRHPRASA 517
Qy      468 LVYKDM 473
Db      518 YWFRDM 523

```

RESULT 10

```

US-10-425-114-70165
; Sequence 70165, Application US/10425114
; Publication No. US20040034888A1
GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ. ID NOS: 2003-04-28
; SEQ ID NO 70165
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73106C01_FLI.pep
US-10-425-114-70165

```

Query Match 37.8%; Score 999; DB 12; Length 527;

Best Local Similarity 42.8%; Pred. No. 8.8e-93;

Matches 208; Conservative 86; Mismatches 148; Indels 44; Gaps 15;

```

Qy      6 ALPNDPEWGATATAYQIEGAVKEGGRGSPSTWDTYCHLEPSTNGANDVACHYHRDED 65
Db      64 SFPKGFVFGTATSAVYQVEGAALSTNGRGPSTWDSFAHVPAGINAGNQNDVAVDQYHRKED 123
Qy      66 FDLITTKGAKAYRFSISMSRIIPLG-GRLDPVNEEGIEFYSKLIIDALLRGTTPWTLVH 124
Db      124 VDLMKSLNPAFYRFSISMSRIIPDGEGK---VNDGVAAYNNLINVLLQGSMTPIYLVH 180
Qy      125 WDLPLQALHDIYGGMLNVEEVLQDFERYARLCFERFGRVONMTITXKPMIQAITYATGS 184
Db      181 YDLPLALEKKYGGMLSAKAMD-L-FTYADFCFTYGRVHKMFTFENPRIVALLGYDTGS 239
Qy      185 NAGRSSINHGSTEGRNATEPMLAGKAQIMSHARAVALVSRDPRPSQKQIGISLNDY 244
Db      240 NPORCT--RCAAGGNSATEPYIVAHNFLAHTAVARVRYTKYQAAKGVIGVLDENWY 297
Qy      245 EPDINSNEPRDKEAERMEFHIGFANPIFLKKDYPSMKKQIGERLPALTPADFAILNA 304
Db      298 EAL-TNSPDDQAAAGARDFHIGFVDP-L-INGHYQIMQDLVKERLPRT-P-EQAKLVK 354
Qy      305 GETDFYGMNYYTSQF--ARHLDGPV-----ETDYLGAIHEHQSNGKSPGEESSGLA 355
Db      355 GSADYIGINEYTSYKMGQVLQVAPSSYSADWQVQVFA-----RNGKPIGPANSK 407

```

```
QY 356 WLRSCPMFRKHLARVGL-----YGRP-IYITENGCPGEBNMTCEAVNDPFR 407
DB 408 WLYIAP-----TGMGCVNVLKKEKGNPTIYITENGMDPG--NLTRDQIRPATVR 458
QY 408 YFDSHLSISKAITODGVVVKGYFAMALLDNLWSDGYGPRFGVFTDYTLKRTPKKSA 467
DB 459 FYRSYIGOLKKAIDQ--GANVAGYFAMSLDNLFWMLAGYSSKFGIYVVDFTLERHPRKASA 517
QY 468 LVLKDM 473
DB 518 YWFRDM 523

RESULT 11
US-10-425-114-71098
; Sequence 71098, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71098
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73086G03_FLI.pep
US-10-425-114-71098

Query Match 37.8%; Score 999; DB 12; Length 527;
Best Local Similarity 42.8%; Pred. No. 8.8e-93;
Matches 208; Conservative 86; Mismatches 148; Indels 44; Gaps 15;

QY 6 ALPNDPEKGFATAYQIEGAVKEGGRGPIWDTYCHLEPSRTNGANGDVACDHYRYDE 65
DB 64 SFPDGFVFGTATSYQVEGASTNGRGPSIWDSPAHVPGNLAGNQGDVAVDQYHRYKED 123
QY 66 FDLTLTKGAKAYRFSLSMSRIIPLG-GRLDPVNBEIGIEFYKLIIDALLRGTIPWTLVH 124
DB 124 VDLKKSINLFDAYRFSISMSRIIPDGBGK--VNPEGVAAYNNLITNLLQCGMTYINLVH 180
QY 125 WDLPOALHRYGGMVNEEVOYLDPERYARLCFERFGDRVQNMITINXPMIOAIYGYATGS 184
DB 181 YDLPLALEKKYGMVLSAMADL-FTDYADFCFTYGRVKKMFNFNPRIVALLGYTGTGS 239
QY 185 NAPERSS--INKESTEGNTATEPMLAGKAOIMSHAANAAYVSRDFRPSQGOIGISLNDY 244
DB 240 NPPRCT--RCAGAGNSATEPEYIVAHNPLLAHATAVARYRTKYOAQKGVGIVLDENWY 297
QY 245 EPMDSNEPRDKEAERMEFHIGWFANPIFLKDYPPSMKKOLGERLPALTPADFAIL 304
DB 298 EAL-TNSPDDQAAQARADRHIGFVDP--INGHYPOIMODLVKERLPRFTP--EQAKLVK 354
QY 305 GETDFYGMNYYTSQF--ARHLDPV-----ETDYLGAIHENKDKSPVGEESGLA 355
DB 355 GSADYIIGINEYTSYMGQKLVQAPSSYSADMQVQVFA-----RNKPIGPQANSK 407
QY 356 WLRSCPMFRKHLARVGL-----YGRP-IYITENGCPGEBNMTCEAVNDPFR 407
DB 408 WLYIAP-----TGMGCVNVLKKEKGNPTIYITENGMDPG--NLTRDQIRPATVR 458
QY 408 YFDSHLSISKAITODGVVVKGYFAMALLDNLWSDGYGPRFGVFTDYTLKRTPKKSA 467
DB 459 FYRSYIGOLKKAIDQ--GANVAGYFAMSLDNLFWMLAGYSSKFGIYVVDFTLERHPRKASA 517
```

```
DB 459 FYRSYIGOLKKAIDQ--GANVAGYFAMSLDNLFWMLAGYSSKFGIYVVDFTLERHPRKASA 517
QY 468 LVLKDM 473
DB 518 YWFRDM 523

RESULT 12
US-10-424-599-213409
; Sequence 213409, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 213409
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_34732C.1.pep
US-10-424-599-213409

Query Match 37.7%; Score 997; DB 12; Length 523;
Best Local Similarity 44.4%; Pred. No. 1.4e-92;
Matches 212; Conservative 78; Mismatches 169; Indels 18; Gaps 12;

QY 6 ALPNDPEKGFATAYQIEGAVKEGGRGPIWDTYCHLEPSR--TNGANGDVACDHYRYDE 64
DB 43 SFPDGFVFGTATSYQVEGASTNGRGPSIWDTFTHRYEIKIDKNSGVAIDAYHYKE 102
QY 65 FDLTLTKGAKAYRFSLSMSRIIPLG-GRLDPVNBEIGIEFYKLIIDALLRGTIPWTLVH 124
DB 103 DVKIVADMNLDYRFSISMSRIIPKGLSRGINQBGIDYNNLNLVANGIQPLVTLPH 162
QY 125 WDLPOALHRYGGMVNEEVOYLDPERYARLCFERFGDRVQNMITINXPMIOAIYGYATGS 184
DB 163 WDLPOGLEDEYGGFLSPRIYK-DFRDYALCFERFGDRVQNMITINXPMIOAIYGYATGS 221
QY 185 NAPERSS--INKESTEGNTATEPMLAGKAOIMSHAANAAYVSRDFRPSQGOIGISLND 242
DB 222 MARGCSAMVNPCTGDSGTEPEYLVTHYOLLAHAANAAYVYRTKYOVOSOKGIGITLVAN 281
QY 243 YPEPMDSNEPRDKEAERMEFHIGWFANPIFLKDYPPSMKKOLGERLPALTPADFAIL 302
DB 282 WYLPF--SNKKAQOKATERAIDPMFGFMDPL--TSGDYPIKIMSIVATRLPKFTTEGSKLL 339
QY 303 NAGETDFYGMNYYTSQFAR--HLDGPVE--TDYIGALIHENKDKSPVGEESGLAWL 357
DB 340 -IGSFDPIGLANTYSSYVSDADPHLSNARPNVYTD--SLVTEPERDGKPIGIKIASDWL 395
QY 358 RSCPMFRKHLARVGLYKGP--IYITENGCPGEBNMTCEAVNDPFRIFDSHLSI 416
DB 396 YVCPRGIDLLLYTEKYNPNPLIYITENGINEFRBETLSLESLDITDFRIDHYRHLFVL 455
QY 417 SKAITODGVVVKGYFAMALLDNLWSDGYGPRFGVFTDY--TLKRTPKKSAALVLKD 472
DB 456 RSLI-RHGVNVAGYIYWSLFDNFEMSSGIVTAFGMLVDYKNNLRKYNHLSAIIWIKN 511

RESULT 13
US-10-425-114-55483
; Sequence 55483, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
```



```

; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 73128
; SEQ ID NO 55483
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMFLMINSOY028C03_FLI.pep
US-10-425-114-55483

```

```

Query Match      37.7%; Score 997; DB 12; Length 552;
Best Local Similarity 44.4%; Pred. No. 1,5e-92;
Matches 212; Conservative 78; Mismatches 169; Indels 18; Gaps 12;

```

```

QY 6 ALNDPEWGFATAAQTGAVKEGGRGSPIMDTYCHLEPSR-TNGANGDVACDHYHRDYE 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 72 SFPDPDFGAGSSSYQFEGAGANEGRGSLIMDTFTHKYPEKIDQKSGVDALDAHYRKYK 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 DFDLLTKRYGAKARFSLMSRRIPLGGRLDPVNEEGIEFYSKIDALLRRGITPMWTLVH 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 132 DVATVDMDLDSYRFSISMSRILPKGLSRGINOEGIDYVNNMLINELVANGIPLVTLFH 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 WDLPOALHRYGGMVLNVEEYQDPERYARLCFERFGDRVQNMITINXPMIOAIYVATGS 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 192 WDLPOLEDEYGFSLRIRYK-DPRDYAEICFKEFGDRVYKWTLINEPMYSQHGANGG 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 NADGRSS--INKSTGNTATEPMLAGKAQIMSHARAAYVSRDFRPSQKQIGISLNGD 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 251 MARGRCSAMVNPCTGDSGTEPYLVTHYQLAHAAVAAYVYKTKYQVSGKLIGITLVAN 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 YVEPMNSNEPRDKEAERMEFHIGWPNRIFLKQVPESMKKQGLERLPALPADPAIL 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 311 WYLPF-SNTYADQKATERALIDFMGFMDPL-TSGDVPKIMRSLVRRRLPKFTTEOSKLL 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 NAGETPFYGMNYYTSQFAR---HLDDGPVE--TDVLCALIHENKDGSPVGEESGLAWL 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 369 -IGSFPFIGLANTYSSRYASAPHLNARNRMYVTD---SLVTPFERGRKIGIKIASDWL 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 358 RSCPDMPFRKHLARVGLYGRP-IYITENGCPCGEEENMTCEAVNDPFRIRYFDSHLDST 416
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 425 YVCPRGILDLTLVTKERYNNPLIYITENGINERDETLSEESLDTFRIDYHRLHYL 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 417 SKAITDDGVVYKGFAMALLDNLEMSDGYPRRGVFTDY-TTLKRTPKKSALVAKD 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 485 RSAI-RHGVNVKGYIWSLFDNFEWSSGYVRGMLLVQYKNLKRHYKLSAIWIKN 540
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14

```

US-10-424-599-147332
; Sequence 147332, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 285684
; SEQ ID NO 147332
; LENGTH: 503

```

```

; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_10405C.1.pep
US-10-424-599-147332

```

```

Query Match      37.7%; Score 996.5; DB 12; Length 503;
Best Local Similarity 42.6%; Pred. No. 1,5e-92;
Matches 201; Conservative 83; Mismatches 163; Indels 25; Gaps 11;

```

```

QY 8 PNDPEWGFATAAQTGAVKEGGRGSPIMDTYCHLEPSR-TNGANGDVACDHYHRDYE 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 32 PNGFVFGTASAFQYGAVALDEKRGSPVMDTFSHTGKIIDFNSADVAVDQYHRYEEDIQ 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 LTRKYGAKYRFSISMSRRIPLGGRLDPVNEEGIEFYSKIDALLRRGITPMWTLVHMDL 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 92 LMKDMGMDARFSLMSRRIPLGNG--YGOINQACVDHNLKILNMLAKGIEPYVTLVHMDL 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 POALHRYGGMVLNVEEYQDPERYARLCFERFGDRVQNMITINXPMIOAIYVATGSNAP 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 150 POALEMKYSGMLNA-SIMDFATYAETCFQKFGDRVYKWTLINEPMYSQHGANGG 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 GRSSINKH--STGNTATEPMLAGKAQIMSHARAAYVSRDFRPSQKQIGISLNGDYE 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 209 GRCSILHLHFCRAGNSATEPEYIVAHNVLLSHATVADLYRKKYKKGSGSGLGAFDVIWYE 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 PMSNEPRDKEAERMEFHIGWPNRIFLKQVPESMKKQGLERLPALPADPAILNAG 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 269 PL-TNTEBEDIDAAQRADDFQGLWFLDPLMF-GDYPSSMRKRVSRLPKTSQSEALVK-G 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 ETDPFYGMNYYTSQFARHLDGPVPEYDLCALIHENKDGSPVGEESGLAWL 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 326 SLDPFVGINHTTYARD-----NSTNLIGTLHDSIADSAVALTLPNGTKAISERASSIW 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 357 LRSCPDMPFRKHLARVGLYGRP-IYITENGCPCGEEENMTCEAVNDPFRIRYFDSHLDST 415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 381 LTVPOSMSKSLMIYIKQYKGNPVIYITENGMDPNISIFISKDALDEKIRIYHTGLSY 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 416 ISKAITDDGVVYKGFAMALLDNLEMSDGYPRRGVFTDY-TTLKRTPKKS 466
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 441 LLSAI-KDGCNVKGYFVWSLDDNWEWSAGTYSRGLFVDYKONLKRYPKQS 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

```

US-10-424-599-219958
; Sequence 219958, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 285684
; SEQ ID NO 219958
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (11).(510)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_4064C.1.pep
US-10-424-599-219958

```

```

Query Match      37.3%; Score 986; DB 12; Length 510;
Best Local Similarity 44.3%; Pred. No. 1,8e-91;
Matches 213; Conservative 77; Mismatches 167; Indels 24; Gaps 12;

```

```

QY 2 PESIAL-----PWFENGFAIATAAYOIEGAYKEGGRPSIMDTYCHLEPSRTNGA-NGDVA 55
Db 31 PEIASLNRSFPPTGFIPTASASAYQYEGANEGGRPSIMDTFTKYPDXIKORDSGDVA 90
QY 56 CDHHRVDEDEPDLLTXGAKAYRFSLSMSRIIPLGRLDPVNEBEGIEFYSKIDALIRG 115
Db 91 IDSYHRYKEDVGIMKDMNLDAYRPSISMSRIIPKGLSGGINQKIDYNNLINELLANG 150
QY 116 ITPWVTLVHMDLPQALHDRVGMNLVEBVLDFERYARLCFERPGRVQNMVITINXPMIQ 175
Db 151 LKRPVTLFHMDDLPGSLDEBYGFLSPRIVK-DFODYADLCFKERGRVKHWITLINEPMSY 209
QY 176 AIYGIATGSNAPEGRS--INKHSTEGNTATEPWLAKAQIMSHARAVAYSRDPRPSOKG 233
Db 210 SQHGVAATGEMAPGRCSAMNPNCGDSDATEPYLVSHHQILAHAAVAVYKTKYQTFONG 269
QY 234 QIGISLNDYVEPMDSNBPRDKEAERMEFPIGFANPIFLKKDYPSMKKOLGERLPA 293
Db 270 LIGITLVNMVYVPSDNK-LDHKATERAIDFOYGFMDPL-TTGDYPSKMRFLVBARLPK 327
QY 294 LTPADFAILNAGETDFYGMNYTTSQFARHLDPVPEYDYLGAIHENKDKDSPVGEESG 353
Db 328 FTKEQSKLL-IDSFDFIGINTYSSASYAS--DAPOLSNAKISYLTDSLNSXSAS----- 378
QY 354 LAMLRSCEDMFRKHLARVYGLYKRP-IYTENGCPCPGEENMTCEAVNDPFRIRYFDSH 412
Db 379 -NMLVYVPRGRDVLVLYTKKKYNNPLIYTENGINEYDSSLSLBSLDDIYRIDYHRH 437
QY 413 LDISKAITODGVVYKGFANALLDNLWMSDYGPRFGVFTDY-TTLKRTPKKSALVLK 471
Db 438 LFYLOEAL-KRGVNVKGYFAMSLLDNFEMHIGYTVRFGMNFIIDYKNDLKRYSKLSALWFK 496
QY 472 D 472
Db 497 D 497

```

Search completed: March 25, 2004, 16:02:25
 Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2004, 15:55:12 ; Search time 182 Seconds
(without alignments)
2595.659 Million cell updates/sec

Title: US-10-026-140-2
Perfect score: 2644
Sequence: 1 MPESLALPNDWEGFATAY.....KSALVLKDPARQRYVAA 484

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Pending_Patents_AA_Main: *
1: /cgn2_6/prodata/2/paa/US06_PCTUS_COMB.pep.*
2: /cgn2_6/prodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/prodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/prodata/2/paa/US080_COMB.pep.*
5: /cgn2_6/prodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/prodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/prodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/prodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/prodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/prodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/prodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/prodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/prodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/prodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/prodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/prodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/prodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/prodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/prodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/prodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/prodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/prodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/prodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/prodata/2/paa/US099_COMB.pep.*
25: /cgn2_6/prodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/prodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/prodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/prodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/prodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/prodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/prodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/prodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	2642	99.9	484	1	PCT-US02-34674-2	Sequence 2, Appli
2	2642	99.9	484	1	PCT-US02-34764-2	Sequence 2, Appli
3	2642	99.9	484	26	US-10-026-140-2	Sequence 2, Appli
4	1303.5	49.3	476	22	US-09-791-537-52753	Sequence 52753, A
5	1270	48.0	466	22	US-09-791-537-52753	Sequence 52755, A
6	1210	45.8	460	29	US-10-369-493-13405	Sequence 13405, A
7	1210	45.8	490	33	US-60-360-039-13405	Sequence 13405, A
8	1204	45.5	524	18	US-09-417-507-31496	Sequence 31496, A
9	1058.5	40.0	491	30	US-10-424-599-169310	Sequence 169310, A
10	1058.5	40.0	511	28	US-10-219-999-35761	Sequence 35761, A
11	1058.5	40.0	511	30	US-10-425-114-43168	Sequence 43168, A
12	1058.5	40.0	511	33	US-10-425-114A-43168	Sequence 43168, A
13	1058.5	40.0	511	33	US-60-324-109-21315	Sequence 21315, A
14	1038	39.3	521	30	US-10-437-963-139890	Sequence 139890, A
15	1037	39.2	514	19	US-09-513-996A-34109	Sequence 34109, A
16	1037	39.2	516	19	US-09-513-996A-34108	Sequence 34108, A
17	1037	39.2	523	19	US-09-513-996A-34107	Sequence 34107, A
18	1025.5	38.8	534	24	US-09-935-625-12100	Sequence 12100, A
19	1020	38.6	515	30	US-10-437-963-144610	Sequence 144610, A
20	1013	38.3	537	22	US-09-791-537-13983	Sequence 13983, A
21	1006	38.0	510	30	US-10-437-963-202947	Sequence 202947, A
22	1003	37.9	577	22	US-09-791-537-80861	Sequence 80861, A
23	999.5	37.8	515	28	US-10-219-999-43869	Sequence 43869, A
24	999.5	37.8	515	30	US-10-425-114-55300	Sequence 55300, A
25	999.5	37.8	515	30	US-10-425-114A-55300	Sequence 55300, A
26	999.5	37.8	515	33	US-60-324-109-30599	Sequence 30599, A
27	999.5	37.8	516	30	US-10-424-599-243532	Sequence 243532, A
28	999.5	37.8	519	28	US-10-219-999-36333	Sequence 36333, A
29	999.5	37.8	519	28	US-10-219-999-47299	Sequence 47299, A
30	999.5	37.8	519	30	US-10-425-114-48403	Sequence 48403, A
31	999.5	37.8	519	30	US-10-425-114-55444	Sequence 55444, A
32	999.5	37.8	519	30	US-10-425-114A-48403	Sequence 48403, A
33	999.5	37.8	519	33	US-10-425-114A-55444	Sequence 55444, A
34	999.5	37.8	519	33	US-60-312-544-6898	Sequence 6898, A
35	999.5	37.8	519	33	US-60-324-109-31263	Sequence 31263, A
36	999	37.8	502	28	US-10-219-999-48605	Sequence 48605, A
37	999	37.8	502	28	US-10-219-999-61094	Sequence 61094, A
38	999	37.8	502	28	US-10-219-999-61151	Sequence 61151, A
39	999	37.8	502	33	US-60-324-109-27920	Sequence 27920, A
40	999	37.8	502	33	US-60-324-109-29463	Sequence 29463, A
41	999	37.8	502	33	US-60-324-109-29539	Sequence 29539, A
42	999	37.8	524	33	US-60-376-344-4	Sequence 4, Appli
43	999	37.8	527	30	US-10-425-114-56976	Sequence 56976, A
44	999	37.8	527	30	US-10-425-114-70165	Sequence 70165, A
45	999	37.8	527	30	US-10-425-114-71098	Sequence 71098, A

ALIGNMENTS

RESULT 1
PCT-US02-34674-2
; Sequence 2, Application PC/TUS0234674
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGLS Beta-Glucosidase and Nucleic Acids
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: PCT/US02/34674
; CURRENT FILING DATE: 2003-03-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Trichoderma reesei
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(484)
; OTHER INFORMATION: Xaa = Any Amino Acid

PCT-US02-34674-2

Query Match 99.9%; Score 2642; DB 1; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.5e-276;
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPSRTNGANGDVACDHYH 60
DB 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPSRTNGANGDVACDHYH 60
OY 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
DB 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
OY 121 TLHYMDLPOLAHDRYGGMLNVEEYQDLPFRYARLCFERFGDRVQWMTITNXPWIOAIYGY 180
DB 121 TLHYMDLPOLAHDRYGGMLNVEEYQDLPFRYARLCFERFGDRVQWMTITNXPWIOAIYGY 180
OY 181 ATGSNAPGRSSINKHSTEGNTATEPMLAGKAQIMSHARAVAVYSRDRPSQKQIGISLN 240
DB 181 ATGSNAPGRSSINKHSTEGNTATEPMLAGKAQIMSHARAVAVYSRDRPSQKQIGISLN 240
OY 241 GDYIEPWSNEPRDKEAERMEFHIGWFANPIFLKKDYPSMKKQIGERLPALTPADFA 300
DB 241 GDYIEPWSNEPRDKEAERMEFHIGWFANPIFLKKDYPSMKKQIGERLPALTPADFA 300
OY 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYIGAIIHEHOKDGSFVGEESGLAMLRSC 360
DB 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYIGAIIHEHOKDGSFVGEESGLAMLRSC 360
OY 361 PDMEFKHLARVYGLYGRPIYITENGCPGGEENMTCEAVNDPFRIRYFDSHLSISKAI 420
DB 361 PDMEFKHLARVYGLYGRPIYITENGCPGGEENMTCEAVNDPFRIRYFDSHLSISKAI 420
OY 421 TODGVVVKGYFAMALLDNLEMSDGYGRFGVTFTDYTLTKRTPKKSALVILKDMFAAROV 480
DB 421 TODGVVVKGYFAMALLDNLEMSDGYGRFGVTFTDYTLTKRTPKKSALVILKDMFAAROV 480
OY 481 KVAA 484
DB 481 KVAA 484
```

RESULT 2
PCT-US02-34764-2
Sequence 2, Application PC/TUS0234764
GENERAL INFORMATION:
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Goedegebuur, Frits
APPLICANT: Ward, Michael
APPLICANT: Yao, Jian
TITLE OF INVENTION: BGLs Beta-Glucosidase and Nucleic Acids
FILE REFERENCE: GC697
CURRENT FILING DATE: 2003-10-30
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 484
TYPE: PRT
ORGANISM: Trichoderma reesei
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(484)
OTHER INFORMATION: Xaa = Any Amino Acid
PCT-US02-34764-2

Query Match 99.9%; Score 2642; DB 1; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.5e-276;
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPSRTNGANGDVACDHYH 60

```
DB 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPSRTNGANGDVACDHYH 60
OY 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
DB 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
OY 121 TLHYMDLPOLAHDRYGGMLNVEEYQDLPFRYARLCFERFGDRVQWMTITNXPWIOAIYGY 180
DB 121 TLHYMDLPOLAHDRYGGMLNVEEYQDLPFRYARLCFERFGDRVQWMTITNXPWIOAIYGY 180
OY 181 ATGSNAPGRSSINKHSTEGNTATEPMLAGKAQIMSHARAVAVYSRDRPSQKQIGISLN 240
DB 181 ATGSNAPGRSSINKHSTEGNTATEPMLAGKAQIMSHARAVAVYSRDRPSQKQIGISLN 240
OY 241 GDYIEPWSNEPRDKEAERMEFHIGWFANPIFLKKDYPSMKKQIGERLPALTPADFA 300
DB 241 GDYIEPWSNEPRDKEAERMEFHIGWFANPIFLKKDYPSMKKQIGERLPALTPADFA 300
OY 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYIGAIIHEHOKDGSFVGEESGLAMLRSC 360
DB 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYIGAIIHEHOKDGSFVGEESGLAMLRSC 360
OY 361 PDMEFKHLARVYGLYGRPIYITENGCPGGEENMTCEAVNDPFRIRYFDSHLSISKAI 420
DB 361 PDMEFKHLARVYGLYGRPIYITENGCPGGEENMTCEAVNDPFRIRYFDSHLSISKAI 420
OY 421 TODGVVVKGYFAMALLDNLEMSDGYGRFGVTFTDYTLTKRTPKKSALVILKDMFAAROV 480
DB 421 TODGVVVKGYFAMALLDNLEMSDGYGRFGVTFTDYTLTKRTPKKSALVILKDMFAAROV 480
OY 481 KVAA 484
DB 481 KVAA 484
```

RESULT 3
US-10-026-140-2
Sequence 2, Application US/10026140
GENERAL INFORMATION:
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Goedegebuur, Frits
APPLICANT: Ward, Michael
APPLICANT: Yao, Jian
TITLE OF INVENTION: BGLs Beta-Glucosidase and Nucleic Acids
FILE REFERENCE: GC697
CURRENT FILING DATE: 2002-03-26
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 484
TYPE: PRT
ORGANISM: Trichoderma reesei
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(484)
OTHER INFORMATION: Xaa = Any Amino Acid
US-10-026-140-2

Query Match 99.9%; Score 2642; DB 26; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.5e-276;
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPSRTNGANGDVACDHYH 60
DB 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPSRTNGANGDVACDHYH 60
OY 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
DB 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120

QY 121 TLVHMDLPQALHRYGGMVNEVQLDFERYARLCEFERFGDRVQNMWTTINXPMIOAIYGY 180
DB 121 TLVHMDLPQALHRYGGMVNEVQLDFERYARLCEFERFGDRVQNMWTTINXPMIOAIYGY 180
QY 181 ATGSNAFGSSINXKSTEGNTATEPWLAKAQMISHARAVAVYSRDRFSPQKQIGISLN 240
DB 181 ATGSNAFGSSINXKSTEGNTATEPWLAKAQMISHARAVAVYSRDRFSPQKQIGISLN 240
QY 241 GDVYEPMDSEPRDKEAERMEFHIGMFANPIFLKQDYPESMKQOLGEERLPALTPADFA 300
DB 241 GDVYEPMDSEPRDKEAERMEFHIGMFANPIFLKQDYPESMKQOLGEERLPALTPADFA 300
QY 301 ILNAGETDFYGMNYTSQFARHLDPVPEYDYLGAIIHEQENKDGSPVGESEGLAWLRSC 360
DB 301 ILNAGETDFYGMNYTSQFARHLDPVPEYDYLGAIIHEQENKDGSPVGESEGLAWLRSC 360
QY 361 PDMFRKHLARVGLYKGPRIYITTEGCPCEPEENMTCEAVNDPFRIRYFDSHLSISKAI 420
DB 361 PDMFRKHLARVGLYKGPRIYITTEGCPCEPEENMTCEAVNDPFRIRYFDSHLSISKAI 420
QY 421 TQGVVYKGFYFAMALLDNLEMSDGYGPRFGVTFTDYTLTKRTPKKSALVLDMPAARQV 480
DB 421 TQGVVYKGFYFAMALLDNLEMSDGYGPRFGVTFTDYTLTKRTPKKSALVLDMPAARQV 480
QY 481 KVAA 484
DB 481 KVAA 484

RESULT 4
US-09-791-537-52753
; Sequence 52753, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEROF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 52753
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Humicola grisea var
US-09-791-537-52753

Query Match 49.3%; Score 1303.5; DB 22; Length 476;
Best Local Similarity 51.7%; Pred. No. 2.9e-131;
Matches 244; Conservative 79; Mismatches 144; Indels 5; Gaps 5;

QY 5 LALPNDPFWGFATTAQVQIEGAVKEGGRPSIWDTYCHLEPSRTGANGADVACDHYHRYDE 64
DB 1 MSLPDPDKMGFAITAAVQIEGAVNEEDGRPSIWDTFCAIPGKIADSSGGAACDYSKATKE 60
QY 65 DPLLLTKYAKAYRSLSMSRIIPLGRLDPVNEEGIEFYSKLIIDALLRGITFWTLVH 124
DB 61 DIALLEKELGANSYRFSISMSRIIPLGGRNDPINKGIDHYKFPVDDILEGTFPFIPLFH 120
QY 125 WDLPOLHRYGGMVNEVQLDFERYARLCEFERFGDRVQNMWTTINXPMIOAIYGYTNGS 184
DB 121 WDLPDALDKRYGGLNKEEPRADENYARIVFKAI-PKCKMIFTFNEPWSGAILGNTGY 179
QY 185 NABGRSSINXKSTEGNTATEPWLAKAQMISHARAVAVYSRDRFSPQKQIGISLANDY 244
DB 180 FAGHTSDRSKSPVGDASAREPMIVGHNIILAHARAVAVYREDFKPTGGEGITLNGDAT 239
QY 245 EPMDSNEPRDKEAERMEFHIGMFANPIFLKQDYPESMKQOLGEERLPALTPADFAILMA 304
DB 240 LPMDDPDPADIEACDRKIEFAISWADPIYFGK-YPDSMRKQGLDRLEPFTPEEVALVK- 297

QY 305 GETDFYGMNYTSQFARHLDPVPEYDYLGAIIHEQENKDGSPVGESEGLAWLRSCDFMF 364
DB 298 GSNDFYGMNYTANYIKHKGVPEDDFLGNIETLLFYNNKYGDCIGETGSMFRLPRAQGF 357
QY 365 RKHLARVGLYKGPRIYITTEGCPCEPEENMTCEAVNDPFRIRYFDSHLSISKAITOP 423
DB 358 RDLNLMVSKRYGYPKLYITTEGSLKGENMDPLEQVLEDFRKYFNDYVRAAAVAED 417
QY 424 GVVYKGFYFAMALLDNLEMSDGYGPRFGVTFTDYTLTKRTPKKSALVLDMP 474
DB 418 GCNVKGLIAMSLLDNLEMSDGYGPRFGVTFTDYTLTKRTPKKSALVLDMP 469

RESULT 5
US-09-791-537-52755
; Sequence 52755, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEROF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 52755
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Hypocrea jecorina
US-09-791-537-52755

Query Match 48.0%; Score 1270; DB 22; Length 466;
Best Local Similarity 52.7%; Pred. No. 1.2e-127;
Matches 246; Conservative 65; Mismatches 142; Indels 16; Gaps 7;

QY 7 LPLNDFEWGFATTAQVQIEGAVKEGGRPSIWDTYCHLEPSRTGANGADVACDHYHRYDE 66
DB 2 LPLNDFEWGFATTAQVQIEGAVNEEDGRPSIWDTFCAIPGKIADSSGGAACDYSKATKE 61
QY 67 DPLLLTKYAKAYRSLSMSRIIPLGRLDPVNEEGIEFYSKLIIDALLRGITFWTLVH 126
DB 62 ALKSLGAKSRYRFSISMSRIIPGCRGDAVNOAGIDHYKFPVDDILEGTFPFIPLFH 121
QY 127 LPLNDFEWGFATTAQVQIEGAVNEEDGRPSIWDTFCAIPGKIADSSGGAACDYSKATKE 186
DB 122 LPLNDFEWGFATTAQVQIEGAVNEEDGRPSIWDTFCAIPGKIADSSGGAACDYSKATKE 180
QY 187 PGRSSINXKSTEGNTATEPWLAKAQMISHARAVAVYSRDRFSPQKQIGISLANDY 245
DB 181 PGRSSINXKSTEGNTATEPWLAKAQMISHARAVAVYSRDRFSPQKQIGISLANDY 240
QY 246 PWDNEPRDKEAERMEFHIGMFANPIFLKQDYPESMKQOLGEERLPALTPADFAILMA 305
DB 231 PWDNEPRDKEAERMEFHIGMFANPIFLKQDYPESMKQOLGEERLPALTPADFAILMA 304
QY 306 ETPDFYGMNYTSQFARHLDPVPEYDYLGAIIHEQENKDGSPVGESEGLAWLRSCDFMF 365
DB 289 SNDFYGMNYTSQFARHLDPVPEYDYLGAIIHEQENKDGSPVGESEGLAWLRSCDFMF 364
QY 366 KHLARVGLYKGPRIYITTEGCPCEPEENMTCEAVNDPFRIRYFDSHLSISKAITOP 424
DB 349 DFLVWIKSKRYGYPKLYITTEGSLKGENMDPLEQVLEDFRKYFNDYVRAAAVAED 417
QY 425 VVYKGFYFAMALLDNLEMSDGYGPRFGVTFTDYTLTKRTPKKSALVLDMP 474
DB 409 VVYKGFYFAMALLDNLEMSDGYGPRFGVTFTDYTLTKRTPKKSALVLDMP 469

RESULT 6
US-10-369-493-13405
; Sequence 13405, Application US/10369493

```
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 13405
/ LENGTH: 490
/ TYPE: PRT
/ ORGANISM: Aspergillus nidulans
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(490)
/ OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-13405
```

```
Query Match      45.8%; Score 1210; DB 29; Length 490;
Best Local Similarity 47.3%; Pred. No. 4.3e-121;
Matches 230; Conservative 82; Mismatches 154; Indels 20; Gaps 4;
```

```
QY 7 LPNDEWGFATAAVYQIEGAVKEGGRPSIWDITYCHLEPSRTNGANGDVACHYHYRDEDF 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 LRNDFHGYATTAAGVAGAWNKDQKPSITWDTFGHTPKVKYKNSNADVAVRFDYFREDV 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 DLITTYGAKAYRSLWSRRIIPLGRLDPVNEEGIEFYSKLIDALLRRGITPMWTLYHMD 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 ALMKSYGVNAVYRSLWSRRIIPLGADDPVNEGGIKYYQDLVDLNNGITPFTLTFHMD 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 LPQALHRYGGMNVEEVLDFERYARLCFERFGDRVONMTITNXPWIAIYGATGNSNA 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 VPQALHEDRYGGMNDRFIPIDFRYARVCFERLGPVYRWMTITNEGVYSLAGYAGVHA 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 187 PGRSSINKESTEGNTATEPMLAGKAQIMSHARAVALVYSDRPPSOQKQIGISLNGDYEP 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 PARSSRELNEEGSDSTEPIVGHTKLVTHGVSKLYREVPFOQKGTIGITLHGWSSE 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 WDSNEPRDEKAERREMEFHIGFANPIFLKQDYPESMKQIGERLPAITPAFALLNAGE 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 WDEDDPRQDEAERAREFEIAMPDPPLKGTGYDPAKMAQIGDRLPRTTPRESKLV-LGS 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 TDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQNKGSPVGEESGLAMLRSCPDMPFK 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 300 SEFYGMNSYTFVFQHKDTPPDINDHKGNVIVHDNNSKGVSGEESDTPMLRTAPTGMRK 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 HLAARYGLGKPIYITENGCCPCPEENMTCEAVNDPRIRYFDSHLSISKAITQDGYV 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 360 LLMWIMNRYHVPYITENGTAKEGTAFT-PEVLIDFRMRFPBGVYVGLAARAKEDGVD 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 VKGYFAMALLDN-----LEWSDGYGPRFGVTFPTDYTLTKRT--PKKSAL 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 IRSYFAMFTDWMGMFKPPPLESDKHREVAAGYTDRCFTIDFDSPMKTRIKOSAY 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 VLKDMF 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 YLKALF 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 7
US-60-360-039-13405
/ Sequence 13405, Application US/60360039
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Chen, Xianfeng
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Hinkle, Gregory J.
```

```
/ APPLICANT: Slater, Steven C.
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)A
/ CURRENT APPLICATION NUMBER: US/60/360,039
/ CURRENT FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 13405
/ LENGTH: 490
/ TYPE: PRT
/ ORGANISM: Aspergillus nidulans
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(490)
/ OTHER INFORMATION: unsure at all Xaa locations
US-60-360-039-13405
```

```
Query Match      45.8%; Score 1210; DB 33; Length 490;
Best Local Similarity 47.3%; Pred. No. 4.3e-121;
Matches 230; Conservative 82; Mismatches 154; Indels 20; Gaps 4;
```

```
QY 7 LPNDEWGFATAAVYQIEGAVKEGGRPSIWDITYCHLEPSRTNGANGDVACHYHYRDEDF 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 LRNDFHGYATTAAGVAGAWNKDQKPSITWDTFGHTPKVKYKNSNADVAVRFDYFREDV 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 DLITTYGAKAYRSLWSRRIIPLGRLDPVNEEGIEFYSKLIDALLRRGITPMWTLYHMD 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 ALMKSYGVNAVYRSLWSRRIIPLGADDPVNEGGIKYYQDLVDLNNGITPFTLTFHMD 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 LPQALHRYGGMNVEEVLDFERYARLCFERFGDRVONMTITNXPWIAIYGATGNSNA 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 VPQALHEDRYGGMNDRFIPIDFRYARVCFERLGPVYRWMTITNEGVYSLAGYAGVHA 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 187 PGRSSINKESTEGNTATEPMLAGKAQIMSHARAVALVYSDRPPSOQKQIGISLNGDYEP 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 PARSSRELNEEGSDSTEPIVGHTKLVTHGVSKLYREVPFOQKGTIGITLHGWSSE 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 WDSNEPRDEKAERREMEFHIGFANPIFLKQDYPESMKQIGERLPAITPAFALLNAGE 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 WDEDDPRQDEAERAREFEIAMPDPPLKGTGYDPAKMAQIGDRLPRTTPRESKLV-LGS 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 TDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQNKGSPVGEESGLAMLRSCPDMPFK 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 300 SEFYGMNSYTFVFQHKDTPPDINDHKGNVIVHDNNSKGVSGEESDTPMLRTAPTGMRK 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 HLAARYGLGKPIYITENGCCPCPEENMTCEAVNDPRIRYFDSHLSISKAITQDGYV 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 360 LLMWIMNRYHVPYITENGTAKEGTAFT-PEVLIDFRMRFPBGVYVGLAARAKEDGVD 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 VKGYFAMALLDN-----LEWSDGYGPRFGVTFPTDYTLTKRT--PKKSAL 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 IRSYFAMFTDWMGMFKPPPLESDKHREVAAGYTDRCFTIDFDSPMKTRIKOSAY 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 VLKDMF 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 YLKALF 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 8
US-09-417-507-31496
/ Sequence 31496, Application US/09417507
/ GENERAL INFORMATION:
/ APPLICANT: KEITH G. WEINSTOCK ET AL.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
/ FILE REFERENCE: PATH99-10
/ CURRENT APPLICATION NUMBER: US/09/417,507
/ CURRENT FILING DATE: 1999-10-14
/ NUMBER OF SEQ ID NOS: 44312
/ SEQ ID NO 31496
/ LENGTH: 524
/ TYPE: PRT
/ ORGANISM: A.fumigatus
```

```

FEATURES:
; NAME/KEY: UNSURE
; LOCATION: (521)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknc
US-09-417-507-31496

```

```

Query Match      45.5%; Score 1204; DB 18; Length 524;
Best Local Similarity 46.9%; Pred. No. 2.2e-120;
Matches 229; Conservative 82; Mismatches 155; Indels 22; Gaps 5;

```

```

QY 7 LPNDFEKGFTAAVQIEGAVKEGGRGSPISWDYCHLEPSTNGANGVACDHYHRDEDF 66
DB 20 LRDFDFHGVATAAQVIEGAVKDKGQSIWDTFHATGKAVDSTGDAAVRSYDLKEDV 79
QY 67 DLTLYGAKAYRFSLSRSRIIPLGRLDPVNEEGIEFSKLDALLRSGITPWTLYHW 126
DB 80 ALMKSYGVNAYRFSLSRSRIIPLGCCDDLVNKEGIEFSKLVDBLNLNGITPVTLYHMD 139
QY 127 LPQALHDRVGMVLANVEVQLDFERYARLCEFRFGDRVQNMWITINXPMWQAIYGYATGSNA 186
DB 140 TPGSLDERYGCMVNOGEKFPVDFVNVYAVCFERLGDRIKHWITNEPBGVYTLAGVAAQVHA 159
QY 187 PGRSSINKESTEGNTATEPMLAGKQIMSHARAVAVSRDPFSQKQIGISLNGDYEP 246
DB 200 PGRSSIFDRNEEGSSTSEPTIYVATTELVAHGVSRLYKQEFQPCQCTIGITLHGNWSEP 259
QY 247 WDSNEPRDKEAARMEFHIQWFPANPIFLKKDYPSMKQLGSRPLALTPADFAILLAGE 306
DB 260 WDAADPLDQAAAEAREFEIAMPADPLKTDYFASHRAQLGDLPTFTBESKLV-LGS 318
QY 307 TDFYGMVNYTSQFARHLDPVETDYLGAITHQENKGSFVGEESGLAWLRSCPDWFRK 366
DB 319 SEFYGMNYSYTFYVAKHTTADINDHKANVEIHDFNKGIPRGESDTEWLRAPWGRK 378
QY 367 HLAHVYGLGKPIYITENGCPCGEENMTCEAVNDPFRIRYPSHLD-SISKAITDDGV 425
DB 379 LLMWIMSRQMPPIYVTENGTTAKG-ETAPSPSVLNDQFRIRFEGYGMALARVAKEDGI 437
QY 426 VVVGYPFAMLLDN-----LEWSDGYGPRFGVFTDYTLTKRT--PKKS 466
DB 438 DISYFAWMTTDMWGFPRFLSSKCAEDGTWMAAGTIDRPGCTIDFDSPEKTRYPKOS 497
QY 467 ALVLKDMF 474
DB 498 AYLIDNLF 505

```

RESULT 9

```

US-10-424-599-169310
; Sequence 169310, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 169310
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123901C.1.pep
US-10-424-599-169310

```

```

Query Match      40.0%; Score 1058.5; DB 30; Length 491;
Best Local Similarity 46.7%; Pred. No. 1.2e-104;
Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

```

```

QY 8 PNDFEWGFATAVQIEGAVKEGGRGSPISWDYCHLEPSTNGANGVACDHYHRDEDF 67
DB 25 PNFIFGVATSAVQIEGACKEGGRGSPISWDAPFHTGKILDKSNGDAVVAHHRVMDID 84
QY 68 LLTLYGAKAYRFSLSRSRIIP--LGRRLDPVNEEGIEFSKLDALLRSGITPWTLYHW 125
DB 85 LIAKLGFDARFISRSRIIPDGLGTR--INDEGITFNNITINGLERIQIOPVTLYHW 141
QY 126 DLQALHDRVGMVLANVEVQLDFERYARLCEFRFGDRVQNMWITINXPMWQAIYGYATGSN 185
DB 142 DPLHLHESGGLNKOILFY-FAVYADTCFASFGGRVKNWITINERLQAVANGVYAI 200
QY 186 AGRSSINKESTEGNTATEPMLAGKQIMSHARAVAVSRDPFSQKQIGISLNGDYEP 245
DB 201 AGRRE-----NSLIEPYLAHHQIILAHAAVAVSYNSKYKXQKQGVFVDCEMAE 252
QY 246 PMSNEPRDKEAARMEFHIQWFPANPIFLKKDYPSMKQLGSRPLALTPAPAI-LNA 304
DB 253 A-NSDXIEDKSAARLDFQLGWFLHPLLY-GDYPEVMERLDDQLPKFSEEDKILLNA 310
QY 305 GETDFYGMVNYTSQFARHLDPVETDYLGAITHQ-----ENKDGSPVGEESGLAWLRSCP 361
DB 311 --LDFIGLHNYTRFISHYTECAEENHY-KVQEMERIYEWEGQAIGERAASEMILYVP 367
QY 362 DMFRKHLARYGLGKPIYITENGCPCGEENMTCEAVNDPFRIRYPSHLDISKAIT 421
DB 368 WGRKILNLYVSQYATPIFVTEGMDDEDNDLPLHEMLDKLRYRFKGYLASVAQAI- 426
QY 422 QDGVVYKGYFAMLLDNLEWSDGYGPRFGVFTDYTL-LKRTKKA 467
DB 427 KQDADVRYGFAMSLDNLEWMAQGYTRFGILVYVDYKNGLSRHKSSA 473

```

RESULT 10

```

US-10-219-999-35761
; Sequence 35761, Application US/10219999
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Dingdong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: cDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52722)C
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/324,109
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/312,544
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 63520
; SEQ ID NO 35761
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Glycine max
US-10-219-999-35761

```

```

Query Match      40.0%; Score 1058.5; DB 28; Length 511;
Best Local Similarity 46.7%; Pred. No. 1.3e-104;
Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

```

```

QY 8 PNDFEWGFATAVQIEGAVKEGGRGSPISWDYCHLEPSTNGANGVACDHYHRDEDF 67
DB 45 PNFIFGVATSAVQIEGACKEGGRGSPISWDAPFHTGKILDKSNGDAVVAHHRVMDID 104
QY 68 LLTLYGAKAYRFSLSRSRIIP--LGRRLDPVNEEGIEFSKLDALLRSGITPWTLYHW 125
DB 105 LIAKLGFDARFISRSRIIPDGLGTR--INDEGITFNNITINGLERIQIOPVTLYHW 161
QY 126 DLQALHDRVGMVLANVEVQLDFERYARLCEFRFGDRVQNMWITINXPMWQAIYGYATGSN 185
DB 162 DPLHLHESGGLNKOILFY-FAVYADTCFASFGGRVKNWITINERLQAVANGVYAI 220

```



```

Oy 186 APERSSINKSTEGNATEMTAGKQOIMSHAAVAVSDRPREQKQIGISLNGYE 245
Oy 221 APERRE-----NSLIERPLAAHQILAHAAVSTIRSKYKQKQGVGVVDCWEAE 272
Oy 246 PMDSNEPRKEAEREMERPHIGWAFNPIFLKKDYPSMKQIGERLPALTPADPAI-IMA 304
Db 273 A-NSDKIEDKSAARRLDFQLGFWLHPLEY- GDYEPVNRERLGDLPFSEEDKKITLINA 330
Oy 305 GETDFGMYNYTOSQFARHLDDGVPEDYDYGALTHEO--ENKDGSPVESEGLMATESCP 361
Oy 331 --LDFIGLHNHYTRFISHYTECAEENHY-KYQEMERIVEMGGALIEKASERLYVP 387
Oy 362 DMFRKHLARVYGLYKPIYITENGCCPCEENMTCEAVNDFRIRYDSDHLSISKAIT 421
Db 388 WGLRKILINYSQKYATPIPIFTEGMDDEBNDNMLPHEMLDKTLRVRYFKGYLAAVQAQAI- 446
Oy 422 ODEGVVYGVYPAWALLDNLEMSDGYGRPFVETDYTTL-LKDPKPKSA 467
Db 447 KDGADYRGGYFAMSLNLFEMAGCYTKRFGLVYVDYKNGLSRHPKSSA 493

```

```

RESULT 11
US-10-425-114-43168
: Sequence 43168: Application US/10425114
: GENERAL INFORMATION:
: APPLICANT: Liu, Jingdong
: APPLICANT: Zhou, Yihua
: APPLICANT: Kovalic, David K.
: APPLICANT: Screen, Steven E
: APPLICANT: Tabaska, Jack E
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(5313)B
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 73128
: SEQ ID NO 43168
: LENGTH: 511
: TYPE: PR1
: ORGANISM: Glycine max
: FEATURE:
: OTHER INFORMATION: Clone ID: 700748475_F11.pep
: US-10-425-114-43168

```

[illegible]

```

QY 362 DMRFKHLAVYGLYKGPITYITENGSGRCGEENMCEEAIVDPPIRIPDLSLISAKIT 442
Db 388 WGIARLKILNVYSGKRAPIPIFVTENGDDDEDNDNLPLHEMLDIDKLAVRIFKGLIASVAQI- 446
QY 422 QDGVVYKGFYFANMLDNLDEMSDGYGPRRGVTFDTYT- LKRTPKKSA 467
Db 447 KQDADVGRGYFAWSLLDNFEWAGYGYKRGGLAVVYDKGLSHHPKSSA 493

```

```

RESULT 12
US-10-425-114A-43168
; Sequence 43168, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425.114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43168
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700748475_F11.psp
US-10-425-114A-43168

```

[illegible]

RESULT 13
US-60-324-109-21315


```

; Sequence 21315, Application US/60324109
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jindong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(5726)B
; CURRENT APPLICATION NUMBER: US/60/324,109
; CURRENT FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 3196
; SEQ ID NO 21315
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; US-60-324-109-21315

Query Match      40.0%; Score 1058.5; DB 33; Length 511;
Best Local Similarity 46.7%; Pred. No. 1.3e-104;
Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

QY 8 PNDPENGATAYQIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDEDF 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 45 PPFIFGVATSAQIIGACKEGGRPSINDAFHTGSKLIDKSNQDVAVNHHRYMEDID 104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 LITKGAAYRPSLSRSRIIP--LGRRLDPVNEEGIEFYSKIDALLRGITPWTLYHW 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 105 LIAKLGFDAYRFISMSRIFFPDGLGTYK--INDEGITFYNNIINGLLERGIQPYVTL 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 DLFOALHDYRGMLNVEVQDLDFERYARLCEFERGDRVQNVITINXPMIOAIYGATGSN 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 162 DLFOALHDSMGMLNKOILEY--FAVYADTCFASFGVRKXMITINBPLOAVANGYDVALF 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 AGRSSINHGTEGNATPEPMLAGKAQIMSHARAVAVYSHDFRPSQKQIGISLNGDY 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 221 AGGRRE-----NLIEPYLAHHQIILAHAAVSVISRYKSKDKKGGVGFVDDCEMAE 272
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 PNDSEPRDKEAERMEFHIGFANPILFKKDYPSMKKQIGERLPALTPADPAI-LNA 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 273 A-NSDIEBKSAARRLDFOLGFWLHPLY--GDYPEVMRERLGDQPKFESEDKILLNA 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 305 GEDDFGNNYTSQFARHLDGVPETDYLGAIHEHQ-----ENKDSVGEESGLAMLRSCP 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 331 --LDFGLNHYTSRFLSHVTECAEENHYX-KVQEMERIVEMEGGOAIGEKAASEMLYVP 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 362 DMFRKHLAVYGLYGRPIYITENGCCPCGSENMTCSEAVNDPRIRYFDSHLSKAIT 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 388 WGRKRLINVSOKYATPIFYTEMGMDDEDNDNLPLEHMLDDKLRVYFKGYLSVAQAI- 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 422 QGVVVKGFAMALLDNLEWSDGYGRPGVYFTDYTT-LKRTPKKA 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 447 KGGADVGYFAMSLDNFEMWAGYTKRFGIVYDYKNGKLSRHKSSA 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-10-437-963-199890
; Sequence 199890, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/437,963

```

```

; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 199890
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_95413C.1.pcp
; US-10-437-963-199890

Query Match      39.3%; Score 1038; DB 30; Length 521;
Best Local Similarity 44.7%; Pred. No. 2.2e-102;
Matches 212; Conservative 80; Mismatches 158; Indels 24; Gaps 10;

QY 6 ALPNDPENGATAYQIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDEDF 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 49 SPEGFVFGTASAAVSDGKEDGQTTWDTFATFGKITDPSNADVADVDYHREED 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 FDLITRYGAKAYRPSLSRSRIIP--LGRRLDPVNEEGIEFYSKIDALLRGITPWTLYHW 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 IQLMADMGDAYRFISMSRIYENG--VQVNOAGIDHNKILIDALLANGIQPYVTL 166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 DLFOALHDYRGMLNVEVQDLDFERYARLCEFERGDRVQNVITINXPMIOAIYGATGSN 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 167 DLFOALHEDYKGLDRIQIVD--DFAAYAEFCFERGDRVQNVITINBPHTVAIOGYDAGLQ 225
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 AGRSSINHG--STEGNTATEPMLAGKAQIMSHARAVAVYSHDFRPSQKQIGISLNGDY 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 226 APCRCSVLHLHYCKAANSCTEPYVAHFTLAAALASIRTKYKATQNGQLAIADVMW 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 YEPWDSNEPRDKEAERMEFHIGFANPILFKKDYPSMKKQIGERLPALTPADPAI-LN 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 286 FEFM-SNTTIDIEAAGRAQEFOLGWFADPFF--GDYPMTRARVGRSLPRFT-ADEAAV 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 AGEDDFGNNYTSQFARHLDGVPETDYLGAIHEHQ-----ENKDSVGEESGL 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 343 KGALEDVGINHYYTYTRH-----NNTNIIIGTLNNTLADTGVSILPFGKXKPIDGRANS 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 AMWRSCPMFRKHLAVYGLYGRK--PIYITENGCCPCGSENMTCSEAVNDPRIRYFDSHL 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 398 IWLIVYPRGRSLNMYKERYNSPVYITENGDDSNPSTISKDLKOSKRIKYHNDYL 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 414 DSISKAITDDGVVVKGFAMALLDNLEWSDGYGRPGVYFTDY-TLTKTPKKS 466
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 458 TNLAASTKDCGVRYFAMSLDNFEMWAGYTKRFGIVYDYKNDLKRYPKKS 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-09-513-996A-34109
; Sequence 34109, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. Alexandrov et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 34109
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..514
; OTHER INFORMATION: any n or Xaa = unknown
; OTHER INFORMATION: Location 1..514 / Ceres Seq. ID 1715035
; US-09-513-996A-34109

Query Match      39.2%; Score 1037; DB 19; Length 514;
Best Local Similarity 44.1%; Pred. No. 2.8e-102;
Matches 209; Conservative 85; Mismatches 156; Indels 24; Gaps 10;

```

```

QY 6 ALPNDEWGPATATAYOIEGAVKEGGRPSIWDTYCHLEBPTNGANGDYACDHYHRYDED 65
DB 29 SFRKGFEVFTASSAFQHEGAVKABRGFTIWDTSHTFGKITDPSNADVAVDQYHRYEED 68
QY 66 FDLITTKYGAAYRPSLSMSRIIPLGRLDPVNEGIEFYSKLI DALIRGITPWTLYHW 125
DB 89 VOLMKNMGMDAYRSISMTWRIFPNG--VGHINEAGIDHYNKLIINALKNGIEPYVTLYHW 146
QY 126 DLPOALHDRVGMWLNVEEVLDPERYARLCEERGRDVRONMTITNXPWIOAIYGYATGSN 185
DB 147 DLPOALHDRVGMWLNPIQIIN-DEFAVAEVCORPGRVXKMTITNEPHTALOGYDVGLQ 205
QY 186 AGRSSINKHST--EGWTATEPMLAGKAQIMSHARAVAVYSRDFRPSQKQIGISLNGDY 243
DB 206 APGRCTILFKLTCEGNSSTEPYIVGHNVILTHATVSDIYRKKYKAKOGGSLGIAFDVW 265
QY 244 YEPWDSNEPRDKAERMEFHIGMFANPIFLKDYEPESMKKOLGERLPALT PADFALN 303
DB 266 FEP-ESNKTEDIEAQRADQFOLGMFLDPLMF-GDYPSMSRVSRLPVFTGSQSSLVK 323
QY 304 AGEUDFYGMNYYTSGFARHLDPVPETDYLGAIHHEQENKDG-----SPVGEESGL 354
DB 324 -GSLDFVGINHYYTYARN-----NATNLIGTLHDAVSDGYTLTPFKGLSTIGDRASS 377
QY 355 AMLRSCPDMPFRKHLARVYGLYK--PIYITENGCCPGSEENMTCEBAVNDPFRIRYFDSHL 413
DB 378 IWLIVPRGMSLNMNVIKHYGNPVPFITENGMDPNILISRKDALMDAKRIKYHHDYL 437
QY 414 DSISKAITODGVVYGYFAMALLDLEWSDGYPFVFTDY--TTLKRTPKS 466
DB 438 SSIQASIKEDGCVKGYFVWSLIDNWEWAAGISRFGLYFDYRDLKRYPKDS 491

```

Search completed: March 25, 2004, 16:00:59
 JOD time : 184 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2004, 15:55:42 ; Search time 15 Seconds
(without alignments)
540.737 Million cell updates/sec

Title: US-10-026-140-2
Perfect score: 2644
Sequence: 1 MPESLALPNDFFWGFATTAAY.....KSALVLKMPARQVAA 484

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 146766 seqs, 16758364 residues

Total number of hits satisfying chosen parameters: 146766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/PCR_NEW_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922	34.9	571	6	US-10-767-701-46509
2	914.5	34.6	565	6	US-10-767-701-46537
3	865	32.7	567	6	US-10-152-372-254
4	857.5	32.4	1927	6	US-10-767-471-1083
5	826	31.2	438	6	US-10-093-037A-16
6	776	29.3	457	1	PCT-US04-02242-96
7	769	29.1	519	6	US-10-767-701-46260
8	765	28.9	462	1	PCT-US04-02242-94
9	702.5	26.6	478	6	US-10-417-884A-4922
10	633.5	24.0	416	6	US-10-767-701-36966
11	610.5	23.1	325	6	US-10-767-701-44803
12	607	23.0	495	6	US-10-417-884A-5715
13	589.5	22.3	470	6	US-10-417-884A-4517
14	567.5	21.5	378	6	US-10-767-701-37355
15	544.5	20.6	278	6	US-10-767-701-37578
16	540.5	20.4	244	6	US-10-767-701-39149
17	540.5	20.4	472	6	US-10-417-884A-6117
18	537	20.3	253	6	US-10-767-701-38415
19	502	19.0	420	1	PCT-US04-02242-90
20	502	19.0	454	6	US-10-093-037A-20
21	470.5	17.8	421	6	US-10-093-037A-17
22	457.5	17.3	485	7	US-10-417-884A-4797
23	449.5	17.0	480	7	US-60-551-121-10
24	446	16.9	503	6	US-10-417-884A-5013
25	445	16.8	437	1	PCT-US04-02242-104
26	437	16.5	210	6	US-10-767-701-39774

27	434.5	16.4	510	6	US-10-093-037A-22	Sequence 22, Appl
28	432	16.3	503	6	US-10-417-884A-6896	Sequence 6896, Ap
29	423.5	16.0	174	6	US-10-767-701-32266	Sequence 32266, A
30	418.5	15.8	497	6	US-10-417-884A-5379	Sequence 5279, Ap
31	414.5	15.7	509	6	US-10-093-037A-18	Sequence 18, Appl
32	410.5	15.5	511	6	US-10-093-037A-21	Sequence 21, Appl
33	409	15.5	484	1	PCT-US04-02242-68	Sequence 68, Appl
34	401.5	15.2	166	6	US-10-767-701-44337	Sequence 44337, A
35	400.5	15.1	512	6	US-10-093-037A-26	Sequence 26, Appl
36	393.5	14.9	512	6	PCT-US04-02242-114	Sequence 114, Appl
37	391	14.8	152	6	US-10-767-701-62456	Sequence 62456, A
38	376.5	14.2	476	1	PCT-US04-02242-20	Sequence 20, Appl
39	373.5	14.1	481	6	US-10-093-037A-15	Sequence 15, Appl
40	371.5	14.1	481	1	PCT-US04-02242-84	Sequence 84, Appl
41	366	13.8	162	6	US-10-767-701-62527	Sequence 62527, A
42	354	13.4	188	6	US-10-767-701-33603	Sequence 33603, A
43	347.5	13.1	314	6	US-10-767-701-43295	Sequence 43295, A
44	347	13.1	180	6	US-10-767-701-58259	Sequence 58259, A
45	343	13.0	171	6	US-10-767-701-62466	Sequence 62466, A

ALIGNMENTS

RESULT 1
US-10-767-701-46509
Sequence 46509, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OR INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53535)B
CURRENT APPLICATION NUMBER: US/10/767, 701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 46509
LENGTH: 571
TYPE: PRT
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-C54_1.pep
US-10-767-701-46509

Query Match	34.9%	Score 922;	DB 6;	Length 571;
Best Local Similarity	40.9%	Pred. No. 9.7e-78;		
Matches 200;	Conservative 87;	Mismatches 184;	Indels 18;	Gaps 13;
QY	1	MPESLALPNDFFWGFATTAAYQIEGAVKEGGGSIIMDTYCHLEPS-RTNGANGDVACDHY	59	
DB	70	IPKRDWFPSPFTTGATISATQIEGWNEDGSGSTMDHFCHTTPDFPIADSGNDVADSY	129	
QY	60	HRVDEPDFLLTKYGAAYRPSLSMSRIPLGRLDPVNEGIEFYSKLDALRGITPW	119	
DB	130	HLVDEPKLLKEMGMAYRFSISMPRLP-NGTSLDINERGIAYNNLHLDNGIEPY	188	
QY	120	VTLHYMDLPALHDIRYGGMLNVEVQDFERYARLCEREGDRVQVMTITNXPMIAIYG	179	
DB	189	VITFHWTPPALVDVYGGFLD-KRIIKDYDFAGLCFEREGDRVNMWLTENEPTFTCLS	247	
QY	180	YATGSNAPGSSSINKISTE--GMTATPEWLAGAOISHRRAVAVVSRRDPSOKGOIGI	237	
DB	248	YGTGILAPGCSGEMKCPDPTGDSIREPYLVGNHLLAHETVDLYNK-FHREKGRIGL	306	
QY	238	SLNGDYVEPMDSNPRDKAABRRMEFHIGFANPIFLKKDYPSMKKQGERLPALTPA	297	
DB	307	ALNMGVTPYGSF-LDEQAHERCMDYNLGMYLEPV-VRGDYPHSMSSVDRLLPHPTK	364	
QY	298	DFAILNAGEIDPFQMTYTSQFARHLD---GPVETDVLGAIHEQENKDSPVGEES	352	
DB	365	EOKL-VGSDYMGINYSRSPFAKHVDITENFSPELNTHDCAT-EETTPNGNTIGPAT	422	


```

Db 217 PGLKL-----RGTGLVKAHHIIKAAKATWHSYNTTWRSKQOGLVSLNCDWGEPR 267
Qy 247 WDSNEPRDKAARERREPHIGFANPIFLKKDYPESKKOLG-----ERLALTP 236
Db 268 VDISNPKDLKAAERYLOFCIGGFANPIY-AGDYPOVWKDYIGKSAEOGLEMSRLPVFSL 326
Qy 297 ADPAINAGETDYGNNVYTSOPARHLDGVPPEPDYLGAIEHOENKD-----GSPVGEES 352
Db 327 QESGYIK-GTSDPLGLGHFTTRITRYTERNYPSRGP-----SYONDRDLIELVDPKMPDL 379
Qy 353 GLAMLRSCDPMFRKLARVYLG-KRIYITENGCPGGEENMTCEBANDPFRIRYFDS 411
Db 380 GSKMLVSPMGFRLLNFAQTOYGDPRIYMEWG---ASQKHCTQ-LCDENRIQYLGK 434
Qy 412 HLDISKAITQDGVVVKGYFAMALLDNLEWSDGYPFGVTFPTDY--TLKRPKKSALV 469
Db 435 YINEMLKAI-KDGANIKGYSWSLDDKFEWEKGYSDRYGYYVEFNDRNKRPYKASVOY 493
Qy 470 LKDMFAA 476
Db 494 YKKIITA 500

```

RESULT 4

```

US-10-767-471-1083
; Sequence 1083, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1083
; LENGTH: 1927
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-767-471-1083

```

Query Match 32.4%; Score 857.5; DB 6; Length 1927;

Best Local Similarity 40.3%; Pred. No. 4, 9e-71;

Matches 192; Conservative 75; Mismatches 167; Indels 43; Gaps 12;

```

Qy 8 PNDPEWGATTAAYOIGAVKEGGRGSPINDYCHLEPSRTNGANGVACDHYRYDEDFD 67
Db 1378 PEGFWSASAAVOIEGAWRADGKLSIMDTFSHTPLRVENDAIQDVACDSYHKIADLV 1437
Qy 68 LRTKYGAKAYRFSLSRSRIIPLGGRLDPVNVEGIEFYSKIDALLRGITPWTLYHMDL 127
Db 1438 TLONLGVSHYRFSISWSRIIP-DGTRRYINEAGINYYVRLIDTLAASIOPOVTIYHMDL 1496
Qy 128 POALHRYGGMVNEEVOUDFERVYARLCEFRGDRVQNMITINXPMIOAIYGYATGSNAP 187
Db 1497 PQLLOD-VGGMEN-ETIVORFKEYADVLFORLDKFKVITLINEPFIATIQGIGYGTAP 1554
Qy 188 GRSSINKHSTEGNTATEPMLAGCAQIMSHARAVAVYSRDRPQKQOIGISLNGDYEPW 247
Db 1555 GVSN-----RPGTAIFYVGNHNLICAHAAEMHLVNDVYRASQGVISITISSDMAEPR 1606
Qy 248 DSNRPDKAERERMEFHIGFANPIFLKKDYPESKKOLG-----FLPALTPA 297
Db 1607 DPNQDEVEAARYVQFMGWFAPHPIFKNDYNEVMKTRIRDSLAAGLKSRLPEFTES 1666
Qy 298 DPAIILNAGETDYGNNVYTSOPARHLDGVPPEPDYLGAIEHOENKDGSPVGE 353
Db 1667 EKRRIN-GTYDFGFRHNTTVLAVNL-----NYATAISSFDADRVAASIABSWPDSC 1718
Qy 354 LAMLRSCDPMFRKLARVYLG-KRIYITENGCPGGEENMTCEBANDPFRIRYFDSH 412
Db 1719 SPWLKMTVPFGFRRLNMLKEEYNDPPIYVTENGV-----SQREETDLNDTARIYVLT 1772

```

```

Qy 413 LDISKAITQDGVVVKGYFAMALLDNLEWSDGYPFGVTFPTDY--TLKRPKKSAA 467
Db 1773 INEALKAV-QDKVDLRGYTWSAMDNFEMATGESERFGLHFVNYSPSLPRIKASA 1828

```

RESULT 5

```

US-10-093-037A-16
; Sequence 16, Application US/10093037A
; GENERAL INFORMATION:
; APPLICANT: Jay M. Short
; APPLICANT: Bylina, Edward
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Lam, David B.
; TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: 564462001402
; CURRENT APPLICATION NUMBER: US/10/093,037A
; CURRENT FILING DATE: 2002-03-06
; PRIOR FILING DATE: US 09/910,579
; PRIOR FILING DATE: 2001-07-20
; PRIOR FILING DATE: US 09/134,078
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 08/949,026
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: US 60/056,916
; PRIOR FILING DATE: 1996-12-06
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 438
; TYPE: PRF
; ORGANISM: Thermotoga sp.
US-10-093-037A-16

```

Query Match 31.2%; Score 826; DB 6; Length 438;

Best Local Similarity 38.4%; Pred. No. 5, 8e-69;

Matches 183; Conservative 71; Mismatches 162; Indels 60; Gaps 14;

```

Qy 8 PNDPEWGATTAAYOIGAVKEGGRGSPINDYCHLEPSRTNGANGVACDHYRYDEDFD 67
Db 8 PKDFIGRTAAVOIEGANEDEGRGSPINDVSHPTGKTLNGDTGVACDHYRYDEDFD 67
Qy 68 LRTKYGAKAYRFSLSRSRIIPLGGRLDPVNVEGIEFYSKIDALLRGITPWTLYHMDL 127
Db 68 LMKELGLDLYRFSISMPRIIMPDKA--INQKVDFTNVLVDELKNDIIPFVTLYHMDL 124
Qy 128 POALHRYGGMVNEEVOUDFERVYARLCEFRGDRVQNMITINXPMIOAIYGYATGSNAP 187
Db 125 PVALYEK-GGMVNL-PIALIFYRAYATFMENLGDRAKMITLINEPWCSSFSGYTGEHAP 182
Qy 188 GRSSINKHSTEGNTATEPMLAGCAQIMSHARAVAVYSRDRPQKQOIGISLNGDYEPW 247
Db 183 GHQNLQ-----EAIITAAHNLREHGHAVQASREEVK--DDEVGITVVMKIEEG 229
Qy 248 DSNRPDKAERERMEFHIGFANPIFLKKDYPESKKOLGEBLPAITPADPAIILNAGET 307
Db 230 DA-KPESFLVASLVDKFNAMSHDPVFGK-YPEEVALYTEKGLOVLDSMNITST-PI 286
Qy 308 DFGNNVYTSOPA-----RHLDGVPPEPDYLGAIEHOENKDGSPVGESEGLAMV 357
Db 287 DFGVNVYTRTLVFPDMNPNLGSVYQGLDKPTBEMGWEIY-----PQGLFDMLVYL 337
Qy 358 RSCDPMFRKLARVYLGKRIYITENGCPGGE-ENMTCEBANDPFRIRYFDSHLS 416
Db 338 KE-----RYKLPLYITENGAGDPKLEN-----GRVHNDYRIEYLEKFEKA 379
Qy 417 SKAITQDGVVVKGYFAMALLDNLEWSDGYPFGVTFPTDYTLKRPKKSALVLDK 472
Db 380 LEAINAD-VDLKGYFWSLMDNFEMACGYSKFGIITYVDYNTPKRIKDSAMMLKE 434

```

RESULT 6

```

PCT-US04-02242-96
; Sequence 96, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; CURRENT FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 96
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: DOMAIN
; LOCATION: (10) --(449)
; OTHER INFORMATION: Glycoeyl hydrolase family 1
PCT-US04-02242-96

```

```

Query Match      29.3%; Score 776; DB 1; Length 457;
Best Local Similarity 37.4%; Pred. No. 2.8e-64;
Matches 182; Conservative 67; Mismatches 171; Indels 66; Gaps 16;

```

```

QY 6 ALPNDPEWGFATAYOIGAVEGGRGSIWDYCHLEPRTNGANGVACDHYHRYDED 65
DB 12 SFPATFTGVATSAVOIGAAALIGRGPSIMDTFSHTEGKIIDSGNGBVADCHYHRAED 71
QY 66 FDLTKYAKAYRFSLSMSRIIPGRLDPVNEEGIEFYSKIDALLRGITPMWTLTYHW 125
DB 72 VELLASLGVAHYRFSMSRVOPTGS--GANNEAGFDYARLLDALAKGIDAHLLTYHW 129
QY 126 DLPALHRYGWLNVESVOLDFERIARLCFERFGDRVQNMITYNXPWIOAIYVATGSN 185
DB 130 DLPALODE--GGWLN--RATCVHFAAYAAEVARFRGHVNASIATINNEPCITAVLGHGTGF 187
QY 186 ARGSSINXKSTEGNTATPEPLAKQAQ--IMSHARAVAYVSRDPRPSQKQIGISLNGD 242
DB 188 APGAA-----DPAAAVOVSHHLLSHGLMQAMRTYNAQA---KLGIVLN-- 229
QY 243 YEEBWD-----SNEPRDKEAERMEFHIGWFANPIFLKQVPESSMKQIGERLPLATPAD 298
DB 230 ---QMTATPATDSADRELALEYARSVQYMDAIF--KGRYPALALAKHIDALSTFEND 285
QY 299 FALINAGETDFYGMNYTTSOPARHLDPGPVETDYLAGIHEHOKDGS PVGEESGLAMLR 358
DB 286 FIDIKQ--FIDELGVNYVYTRAFMSAETPRKRECKLGV-----NDMGW-E 327
QY 359 SCPDMFRKHARVYGLVG-KPIYITENGSC-----PCGGEENMTCEAVNDPFRIRYDSH 412
DB 328 TYPGCLTELLGLHRETRLPVYITTEGMVAADRPVNGK-----VHDEPREYVRH 379
QY 413 LDISKAITODGVVVKGFAMALDNLMSDGYGRFGVTFDTYTLTKRPPKSATLVKX 472
DB 380 LDAL-RAVVAQGIYRGVYFWSLMDNFENMSGAKRGMIVDYATQGRSFKOSALMYRD 438
QY 473 MEAAQ 478
DB 439 FIAAQ 444

```

```

RESULT 7
US-10-767-701-46260
; Sequence 46260, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

```

```

; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 46260
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAV03-C4596_1.pep
US-10-767-701-46260

```

```

Query Match      29.1%; Score 769; DB 6; Length 519;
Best Local Similarity 38.3%; Pred. No. 1.5e-63;
Matches 179; Conservative 83; Mismatches 181; Indels 24; Gaps 14;

```

```

QY 8 PNDPEWGFATAYOIGAVEGGRGSIWDYCHLEPRTNGANGVACDHYHRYDED 67
DB 34 PAGEVFGVSSAYQVGAVAEDGRKPSIMDTFTHKYS-IDATGVDYADQYHKYKEDV 92
QY 68 LTKYAKAYRFSLSMSRIIPGRLDPVNEEGIEFYSKIDALLRGITPMWTLTYHW 127
DB 93 LHELGVDAYRMSIAMPRLIPDS--RGAVNPKGLEIYNNLIBELSYGIQPHVTIYHPD 150
QY 128 PQLHRYGWLNVESVOLDFERIARLCFERFGDRVQNMITYNXPWIOAIYVATGSNAP 187
DB 151 PQLODEYKGLSPREIE--DYTAAYADVCFKNGDRKVSIVNEPIETIGFDQIILP 209
QY 188 GRSS--INXSTEGNTATPEPLAKQAQIMSHARAVAYVSRDPRPSQKQIGISLNGD 245
DB 210 QRSAPFGGCDGNSNTTPTTAHHLVLAHSAVSLYERKQIQAEGGSIADALLGMYE 269
QY 246 PWDNEPRDKEAERMEFHIGWFANPIFLKQVPESSMKQIGERLPLATPADFALINAG 305
DB 270 P-ATETPDDIAAARNDPHVGMFHM-VMGDYVPEVKVNGSRLPSTNNE-AKRVYG 326
QY 306 EDPFYGMNYTTSOPAR---HDDGPVETDY--GAIEHQ--ENKQSPVGEESGLAM 356
DB 327 SFDVGFNHYVAVYVADLSRLDEKV--RDYADAAYVADMPFLKSRNQFPFGLGLTADP 384
QY 357 LNSCPDMFRKHARVYGLVGK-PITYTENGSCPCGGEENMTCEAVNDPFRIRYDSH 415
DB 385 TITPTPAALKMLKHQVYKKNPAWIHENG--AAQOPDSSGVNTYDDEKRSQYLDYIEA 442
QY 416 ISKAITODGVVVKGFAMALDNLMSDGYGRFGVTFDTYTLTKRT 462
DB 443 TLQSI-RNGSNVQGYFVMSFLDVFEYLFGRYRFGVGVETASTART 488

```

```

RESULT 8
PCT-US04-02242-94
; Sequence 94, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; CURRENT FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 94
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample

```

FEATURE:
NAME/KEY: DOMAIN
LOCATION: (9)...(462)
OTHER INFORMATION: Glycosyl hydrolase family 1
PCT-US04-02242-94

Query Match 28.9%; Score 765; DB 1; Length 462;
Best Local Similarity 36.9%; Pred. No. 3e-63;
Matches 176; Conservative 76; Mismatches 171; Indels 54; Gaps 16;

QY 5 LAIPNDFEWFATPAAYOIEGAVKEGGRPSINPTCYCHLEPSRTNGANGDVACHYHYRDE 64
DB LSPPEQFWGAAAYOVBEAGVHEDGSPVWDFCEKPGAVQGHGAAVACHYHYRRE 69
QY 65 DFLTLTKYGAAYRFSLSWSRIIPLGRLDPVNEBEGIEFYSKLIIDALLRSGITPWTLYH 124
DB DVALMQVGLHAYRLSVCPVPEG--VGQPNKGLDFTSRVLDALLEGITPWTLYH 127
QY 125 WDLPOALHRYGGLVNEVQDLFERVYARLCFERFGDRVQNMWTTINXPMIOAIYGYATGS 184
DB WDYPALAYHR--GGMLNRDSADW--FAEYAGLIADRLSDRVQHFPTQNEPQYIGFGLHKG 185
QY 185 NAGRSISNGHSTEGTATEPMLAGKQIMSHAAVAAYSRDRPQSGQIGISLNGDY 244
DB HAEG-----DTLPMSCVLLAGHSHSLAHGKAVALRA--QAKQQLRGVAPVGMPL 234
QY 245 EPWDSNEPRDKAEARMEFHI-----GWFANPIFLKDYPSMKQLGERLPALTP 236
DB 235 HRP--TDSADVAAKATFTWREKSNMAMMDPFL--GETPAQGLAFPGDVPQVRE 231
QY 297 ADEFAINAGETDFYGMNYYTSQFARHLDPVPEVDYLGAIHEHQNKGSPVEESGLAW 356
DB 292 GDMQLI-AQPLDFPGVNIYSTPVR--ASSAESGF-----EVVPHPTGYPI--TAFNW 339
QY 357 -----LRSCPDMFRKHLAAYVGLYKPIYITTEGCPCEENKTCBAVNDPFRIRYFD 410
DB 340 PITPOLAYGPRPF-----YERYOPRIVITENGLSC--RDVAVADGKXHDARIDFTT 390
QY 411 SHLDSISKAITOGVVVKGYFAMALLDNLEMSDGYGPRFGVTFTDYTLTKRTPKKA 467
DB 391 RYLRHLHRAVA-DGVAVEGTFHWSINDNFEMAGYRERFGLIHVDYETLAKTPKAS 446

RESULT 9

US-10-417-884A-4922
Sequence 4922, Application US/10417884A

GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street

CITY: Waltham
STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/417,884A

FILING DATE: 17-Apr-2003

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532

FILING DATE: 30-Jun-1998

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4922:
SEQUENCE CHARACTERISTICS:
LENGTH: 478 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...478
SEQUENCE DESCRIPTION: SEQ ID NO: 4922:
US-10-417-884A-4922

Query Match 26.6%; Score 702.5; DB 6; Length 478;
Best Local Similarity 34.2%; Pred. No. 2e-57;
Matches 169; Conservative 81; Mismatches 179; Indels 65; Gaps 15;

QY 3 ESALPNDPEWGPATAYOIEGAVKEGGRPSINPTCYCHLEPSRTNGANGDVACHYHYR 62
DB 5 ELTSPKDFLWGSASAAVYVEGAMQEDGSESWDFVRI PGKTFKGTGDLVADYHNR 64
QY 63 DEEDFLTKYGAAYRFSLSWSRIIPLGRLDPVNEBEGIEFYSKLIIDALLRSGITPWTLY 122
DB 65 KEDIALMKQGLKAYRFSVAMTRIPNG--RGVNOAGLAFYERLDELLENDEEMTL 122
QY 123 YHMDLPQALHRYGGLVNEVQDLFERVYARLCFERFGDRVQNMWTTINXPMIOAIYGYAT 182
DB 123 YHMDLPQALQDEYNGWESRQITE-DPTHYAETLFEAFRGKVHMYSLNQNIFTSISGLYL 181
QY 183 GSNAPRSISNGHSTEGTATEPMLAGKQIMSHAAVAAYSRDRPQSGQIGISLNGD 242
DB 182 AAHPG-----VTPKRMVEVNHIANLANNAVINKFHELEMPGKIGSPFAYT 228
QY 243 YEEDPDSNEPRDKAEARMEFHI--GWFANPIFLKDYPSMKKQLGER--LPALTPADPA 300
DB 229 PNYPIDSN--PENVALAENMEDLMANYKMD--VTMGKYPILAAKRFLEKGMARPIEAGDAE 286
QY 301 ILNAGETDFYGMNYYTSQFARHLDPVPEVDYLGAIHEHQNKGSPVEESGLAWLRSC 360
DB 287 LLESARPDDELGINVY--QTATNAFNL--NGVAGAKNATGKGS--SEETGV----- 333
QY 361 PMFRK-----HLARVYGLYKPIYITTEGCPCE--ENMTCE 397
DB 334 PGMYKVENPFWERTNWEIDPEGLRIGLRITSRYPVLTTEGL--GEYDLTATD 390
QY 398 EAVNDPFRIRYFDLSDISKAITOGVVVKGYFAMALLDNLEMSDGYGPRFGVTFTDYT 457
DB 391 KOIHDDYRINTYLOSHKAIKEAIS--DGAVALGCTMSTYDLSLNGLYOKRIGFVVDOD 449
QY 458 ----TLKRTPKKS 466
DB 450 ETQEGSLERYKKS 463

RESULT 10

US-10-767-701-36968
Sequence 36968, Application US/10767701

GENERAL INFORMATION:

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53535)B

```

1      RESULT 12
2      US-10-417-884A-5715
3      : Sequence 5715, Application US/10417884A
4      : GENERAL INFORMATION:
5      : APPLICANT: Lynn A Doucette-Stamm and David Bush
6      : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
7      : ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
8      :
9      : NUMBER OF SEQUENCES: 7310
10     : CORRESPONDENCE ADDRESS:
11     : ADDRESSEE: GENOME THERAPEUTICS CORPORATION
12     : STREET: 100 Beaver Street
13     : CITY: Waltham
14     : STATE: Massachusetts
15     : COUNTRY: USA
16     :
17     : ZIP: 02354
18     :
19     : COMPUTER READABLE FORM:
20     : MEDIUM TYPE: CD-ROM ISO9660
21     : COMPUTER: PC
22     : OPERATING SYSTEM: <Unknown>
23     :
24     : SOFTWARE: ASCII
25     :
26     : CURRENT APPLICATION DATA:
27     : APPLICATION NUMBER: US/10/417,884A
28     : FILING DATE: 17-Apr-2003
29     :
30     : PRIOR APPLICATION DATA:
31     : APPLICATION NUMBER: US/09/107,532
32     : FILING DATE: 30-Jun-1998
33     : APPLICATION NUMBER: 60/085,598
34     : FILING DATE: 14 May 1998
35     : APPLICATION NUMBER: 60/051571
36     : FILING DATE: July 2, 1997
37     :
38     : ATTORNEY/AGENT INFORMATION:
39     : NAME: Ariniello, Pamela Deneke
40     : REGISTRATION NUMBER: 40,489
41     : REFERENCE/DOCKET NUMBER: GTC-012
42     :
43     : TELECOMMUNICATION INFORMATION:
44     : TELEPHONE: (781)893-5007
45     : TELEFAX: (781)893-8277
46     :
47     : INFORMATION FOR SEQ ID NO: 5715:
48     : SEQUENCE CHARACTERISTICS:
49     : LENGTH: 495 amino acids
50     : TYPE: amino acid
51     : TOPOLOGY: linear
52     :
53     : MOLECULE TYPE: Protein
54     :
55     : HYPOTHEITICAL: YES
56     :
57     :
58     :
59     :
60     :
61     :
62     :
63     :
64     :
65     :
66     :
67     :
68     :
69     :
70     :
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :

```



```

; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...495
; SEQUENCE DESCRIPTION: SEQ ID NO: 5715:
US-10-417-884A-5715

Query Match      23.0%; Score 607; DB 6; Length 495;
Best Local Similarity 35.7%; Pred. No. 1,6e-48;
Matches 177; Conservative 56; Mismatches 181; Indels 80; Gaps 24;

QY 7 LPNDFEWGATAAAYQIEGAVKEGGRPSIWDYTCHEPSTNGANGDVACDHYRRYDEDF 66
DB 32 LPKDFIFGATAYQAGATHTDGKGVANDKY--LEDNYWYA--EPASDFHKKPVLD 87
QY 67 DLTITGAKAYRSLSMSRIIPLEGRLDPNREGIERYSKIDALLRGITTPWTLTHMD 126
DB 88 ELAEKYGAVGIRISIMSRIFPTG--YGEVNEKGVETRYHNLFAECHKRHVEPVTLLHFD 145
QY 127 LPQALHRYGGMNVEVOLDFERYARLCPERFGDRVQNMITINXPMQAIYGATGSSNA 186
DB 146 TPEVLHSD--GDFLNRENIR--HFVDYAFCEEPPE--VNYVTTTNEIGPIDGQILVSKFP 202
QY 187 PG-----RSSINRGSTEGTATEPMLAGKAQIMSHARAVALVSRDPRPSQKQIGI--SL 239
DB 203 PGIQYDLAKVFGSHHN-----MMVSHARAVLYK---DKGYKGEIGVYVHAL 245
QY 240 NGDIYEPWDSNERDYEAEARRMEFHIGFANPIFLKKDYPSMK-----KOLGERLPA 293
DB 246 PTXY--PYDPENPADVRAALEEDIINHKFLDATTYLGHVSDKTMGEVNHILANGELD- 302
QY 294 LTPADFAILNAGE--TDFYGMNYYTSQFARHLDPVPTDYLGAIHGHQNKDGPVCEE 351
DB 303 LRPEDFOALEAKDLNDFLGINYMSDWMQAPDG--ETG--IHHGKEK--GSSKTYQI 355
QY 352 SGLAWLRSCPEMRKILAR-----VY--GLYG-----KPIYITENGCPGPG 391
DB 356 KGVG--RIAPD-----YVPRTDWMIIYPEGLYDQIMKVNKDYPNKKIYITENGL---GY 407
QY 392 ENMTCEAVNDPRIRYFDSHLDSISKAITQDGVVVKGYFAMALLDNLEWSDGYGPRFGV 451
DB 408 KGEFVNTVYDDRIYVKKHLEVLSDALA--DGANVKGFIYMSIMDFWSNIGYKRYGL 466
QY 452 TETDYTTLKRTPKKSA 467
DB 467 FYVDFDIOERYPKSA 482

RESULT 13
US-10-417-884A-4517
; Sequence 4517, Application US/10417884A
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSES: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/417, 884A
; FILING DATE: 17-Apr-2003
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/09/107,532
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: CTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4517:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...470
; SEQUENCE DESCRIPTION: SEQ ID NO: 4517:
US-10-417-884A-4517

Query Match      22.3%; Score 589.5; DB 6; Length 470;
Best Local Similarity 32.9%; Pred. No. 6,4e-47;
Matches 168; Conservative 75; Mismatches 161; Indels 107; Gaps 25;

QY 5 LALPNDFEWGATAAAYQIEGAVKEGGRPSIWDYTCHEPSTNGANGDVACDHYRRYDE 64
DB 6 LKLPEDFIFGATAYQAGATHTDGKGVANDKY--LEDNYWYA--EPASDFHKKPVLD 87
QY 65 DFLITGAKAYRSLSMSRIIPLEGRLDPNREGIERYSKIDALLRGITTPWTLTHMD 123
DB 62 DIELGERFGVNGIRISIMSRIFPDGAKPNP--EGIAFYHRVFECKRANTPTVTLH 118
QY 124 HMDLPQALHRYGGMNVEVOLDFERYARLCPERFGDRVQNMITINXPMQAIYGATG 183
DB 119 HFDTPRPLFD--HGFNLNREITEA--FVSVAIFCFHEFKE--VKWSTFNEIYPAVATNOYLLG 175
QY 184 SNAPG--RSSINK-----HSTEGTATEPMLAGKAQIMSHARAVALVSRDPRPSQKQIGI 237
DB 176 VPPPGIKYDFTKIYVACLHN-----MMVAHARVNVYKNEHLR--GEIGV 217
QY 238 --SLANGDYEPWDSNERDYEAEARRMEF-----HIGWFANPIFLKKD----- 278
DB 218 VHSLEFKYAA--TDAPEBKHAFLDDALSIRFLDATTYLGHVSTETLTLALDICEANQA 274
QY 279 ---YPSMKKQIGERLPALTPADFAILNAGE--DFYGMNYYTSQFARHLDPVPTDYLG 333
DB 275 SYHFPPE-----DFVELKKASTDNDYLGINHVOCHPVKAYDGE----- 312
QY 334 GAHGHQNKDGPVCEESGLA-----WLASCPMPFKHILARV---GYLYG 376
DB 313 NALHNHGTEKGSYVKKVIGIRIYKEGIPTRTDML--IYPEGLYDILRLIRISDPHYN 371
QY 377 KPIYITENGCPGCEENMTCEAVNDPRIRYFDSHLDSISKAITQDGVVVKGYFAMALL 436
DB 372 K-IYITENGM---GYDQDFEDGIIMQPRIDYLRVYVLESLSKAITA--GVNVKGYFLMSLM 426
QY 437 DNLEWSDGYGPRFGVTFDTYTTLKRTPKKSA 467
DB 427 DLFSTWNGYNKRYGLFYVDFETQKRYPKESA 457

RESULT 14
US-10-767-701-37355
; Sequence 37355, Application US/10767701
; GENERAL INFORMATION:

```

APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 37355
LENGTH: 378
TYPE: PRT
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-C2562_1.pep
US-10-767-701-37355

Query Match 21.5%; Score 567.5; DB 6; Length 378;
Best Local Similarity 35.5%; Pred. No. 5.4e-45;
Matches 125; Conservative 66; Mismatches 150; Indels 11; Gaps 10;

QY 118 PWTLYHMDLPQALHDIRGYGLNVEVOLDPERYARLCEERFGDRVQNMWITINXPMIOAI 177
DB 4 PFTVTHVDIPELQORVGSWLS-PEIQEDFTYFAELCFKMFGRVAKMATEFNENLMAK 62
QY 178 YGATGSNAPGRSS--INKHSTEGNTATEPWLAKQAQIMSHARAVALVYSRDFRPSQKQI 235
DB 63 IAYFGKPPSPHSCSPFGK-CNSGNSSTEPYLAHNMILAAKTVNIYKKNYKTKOGGSV 121
QY 236 GISLNGDYYPWDSNEPRDKAEERMEFHIGFANPIFLKDYVESMKKOLGERPLAT 295
DB 122 GIVVWRWYEPV-SVYTDHILAVSRAQSFAPWFLDPLFF-GDYPHQMRQILGPMLEFT 179
QY 296 PADFAILNAGDFVGMNNTYTSQFARHLDGPVETD-YLG-AIHENKDKGSPVGEESG 353
DB 180 EGEKQLMK-NQIDFISVNHKTLVYDCVYSLCDDITYAGDALVSESAERNIGIPGKPTP 238
QY 354 IAWLRSCPDMPFRGLARVYGLYK-PIYITENGSCPCGEEHNTCEAVNDPRIRYPSH 412
DB 239 VANNVVPSSMELVWYLTQRYKSIPLYITENGVAQIGNSPPTBELINDTERSSYINDY 298
QY 413 LDISKAITODGVVYKGFYFAMALDNLWSDGGRFGVTFDPTTLKRTPK 464
DB 299 LTVLSFAL-RKGADVGRYFVMSLMDNFEMLSGTTIKGYHYDFKSLKRTPK 349

RESULT 15

US-10-767-701-37578
Sequence 37578, Application US/10767701

GENERAL INFORMATION:
APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement

FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701

CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128

SEQ ID NO 37578
LENGTH: 278

TYPE: PRT
ORGANISM: Sorghum bicolor

FEATURE:
NAME/KEY: unsure

LOCATION: (1) (278)
OTHER INFORMATION: unsure at all Xaa locations

OTHER INFORMATION: Clone ID: SORBI-28MAY03-C31549_1.pep
US-10-767-701-37578

Query Match 20.6%; Score 544.5; DB 6; Length 278;
Best Local Similarity 44.7%; Pred. No. 4.9e-43;

Matches 109; Conservative 45; Mismatches 79; Indels 11; Gaps 5;
QY 24 GAVEGGRGSPIMDTYCHLEPSPRTNGANGVACDHYHRDDEBDLTKYGAAXYRFSLSM 83
DB 46 GAREGGKSDSIMDVFTDDERVLDKSMAEIAVDHYRKYEDIELMASLGFSAKRSISM 105
QY 84 SRITPLGGRLDVNEEGIEFYSKLIDALLRRGITPWTLYHMDLPQALHDIRGYGLNVEE 143
DB 106 ARIFP-DLGEKVEQVAFINDLINMISKGIEPATLYHMDLPNNLQKTLGGWISDKI 164
QY 144 VQDPERYARLCEERFGDRVQNMWITINXPMIOAIYGAATGSNAPGRSSINKHSTEGNTAT 203
DB 165 VEY-FALVAACFANFGDRVAKWITINEPOTAIYGIIFAPG-----GCQGETA- 215
QY 204 EPMLAGKAQIMSHARAVALVYSRDFRPSQKQIGISLNGDYYPWDSNEPRDKAEERME 263
DB 216 RCTLAHHQILAAHAAVAVYRRKFAAGGGEVGVDCEMAEFP-SKAEQIQAQRRID 274

Search completed: March 25, 2004, 16:01:26
Job time: 16 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 20:57:07 ; Search time 5341.5 Seconds
(without alignments)
11130.890 Million cell updates/sec

Title: US-10-026-140-1
Perfect score: 1991
Sequence: 1 agcagcgtcgcacacagcag.....ctttaaaaaaaaaaaaaa 1991

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: 1: em_gstba:*
2: em_gsthum:*
3: em_gstln:*
4: em_gstnu:*
5: em_gstov:*
6: em_gstpl:*
7: em_gstro:*
8: em_hlc:*
9: gb_gst1:*
10: gb_gst2:*
11: gb_hlc:*
12: gb_gst3:*
13: gb_gst4:*
14: gb_gst5:*
15: em_gstfun:*
16: em_gstcom:*
17: em_gsthum:*
18: em_gstln:*
19: em_gstpl:*
20: em_gstro:*
21: em_gstfun:*
22: em_gstman:*
23: em_gstmus:*
24: em_gstpro:*
25: em_gstrod:*
26: em_gstphg:*
27: em_gstvrl:*
28: gb_gst1:*
29: gb_gst2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	702.2	35.3	773	14	CB904679 trico39xe
2	701	35.2	703	14	CF875727 trico39xe
3	461.2	23.2	584	14	CF884351 trico39xe
4	387	19.4	1994	9	AA415086 MG0020 RC

5	201	10.1	2031	11	AY109400
6	195.2	9.8	788	14	CB677158 OSJNE013N
7	193	9.7	786	14	CB673495 OSJNE08A
8	190.4	9.6	806	14	CB642881 OSJNE03E
9	179.2	9.0	889	14	CD438244 EL01N0511
10	175.6	8.8	480	14	CB688082 CBST-23-E
11	175	8.8	1182	11	AY106991 Zea mays
12	171.8	8.6	662	14	CF880605 trico082xh
13	171.8	8.6	752	14	CB907493 trico82xh
14	169.8	8.5	650	10	BE77259 MY-26-C-1
15	162.6	8.2	613	9	AJ436490 AJ436490
16	158.6	8.0	860	14	CF666831 R1CNT1_26
17	157.8	7.9	610	13	BO110523 VD0105D5
18	157	7.9	825	14	CK200642 FGAS00915
19	156.2	7.8	1126	14	CK209455 FGAS02132
20	155.8	7.8	708	13	CA145286 SCGRST206
21	155.6	7.8	673	14	CD482402 atc01-26m
22	155.6	7.8	674	14	CD458246 FG08_06h0
23	154	7.7	786	14	CB634193 OSIEB131
24	153.8	7.7	1000	14	CF933596 TREST-B08
25	153.4	7.7	792	14	CB633739 OSIEB12M
26	153.4	7.7	808	14	CB633785 OSIEB12M
27	153.2	7.7	739	13	CA139881 SCEZRT201
28	153	7.7	1126	14	CK161716 FGAS01428
29	152.4	7.7	788	14	CF554428 G9M12 Inf
30	151.8	7.6	728	14	CF473575 RTW2 3 H
31	151.6	7.6	723	12	BM605139 170006870
32	151.6	7.6	766	14	CB643857 OSJNEB04M
33	151.6	7.6	821	14	CB634605 OSIEB14D
34	151.2	7.6	564	14	CB641046 OSJNEA17J
35	151.2	7.6	636	9	AV913799 AV913799
36	151	7.6	676	12	BM621029 170006874
37	151	7.6	805	14	CB677500 OSJNEA14I
38	150.8	7.6	646	12	BJ235347 BJ235347
39	150.8	7.6	839	14	CK201813 FGAS01033
40	150.6	7.6	829	14	CB663657 OSJNEB080
41	150	7.5	592	14	CA193361 SCRLAD104
42	150	7.5	760	14	CB643189 OSJNEB03L
43	150	7.5	794	14	CB654045 OSJNEC05K
44	150	7.5	806	14	CB674932 OSJNE10E
45	150	7.5	826	14	CB648777 OSJNEB12P

ALIGNMENTS

RESULT 1
CB904679 773 bp mRNA linear EST 02-JUL-2003
LOCUS trico39xe13 T. reesei mycelial culture, Version 3 apr11 Hypocrea
DEFINITION jecorina cDNA clone trico39xe13. mRNA sequence.
ACCESSION CB904679
VERSION CB904679.1 GI:30119337
KEYWORDS EST.

SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Bukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE

1 (bases 1 to 773)
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)

TITLE

JOURNAL
MEDLINE
PUBMED

22803314
12788920

Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817

COMMENT

Email: Pforeman@genencor.com
Seq primer: LR-F1 primer.
Location/Qualifiers

RESULT 2
CF875727 703 bp mRNA linear EST 31-OCT-2003
LOCUS tric039xe13.b1 T. reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone tric039xe13, mRNA sequence.

FEATURES
source
1..773
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="Qme6"
/db_xref="taxon:51453"
/clone="tric039xe13"
/dev_stage="mycelia"
/clone_lib="T. reesei mycelial culture, Version 3 april"
/note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 35.3%; Score 702.2; DB 14; Length 773;
Best Local Similarity 92.8%; Pred. No. 4.8e-114; Indels 0; Gaps 0;
Matches 707; Conservative 1; Mismatches 54;

340 GGATCCCGCTCAACGAGGAGGAATTGAGTTTACAGCAAACTGATGACGCTTGTGAG 399
12 GNNNCTGCTGCTGTAANNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 71
400 GCGGGGATACAGCCTTGGGTGATCTTTTACACCTGGGATCTGCTCAAGGCGTTACGA 459
72 GCGGGGATACAGCCTTGGGTGATCTTTTACACCTGGGATCTGCTCAAGGCGTTACGA 459
460 TCGGATGAGGCTGCTCAACGATGAGGAGGTCAGCTGAGCTTTGAGCGGTATGCGAG 519
132 TCGGATGAGGCTGCTCAACGATGAGGAGGTCAGCTGAGCTTTGAGCGGTATGCGAG 519
520 GTTGTGCTTTGAAAGTTTGGGAGCGAGTCCAGAACTGATCACCATCAACGAMCCCTG 579
192 GTTGTGCTTTGAAAGTTTGGGAGCGAGTCCAGAACTGATCACCATCAACGAMCCCTG 579
580 GATTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 639
252 GATTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 639
640 CAAGCACTCCACCGAGGAGGCAACCTGCACTGAGCGGTGCTGCTGGAAGGCCCAAT 699
312 CAAGCACTCCACCGAGGAGGCAACCTGCACTGAGCGGTGCTGCTGGAAGGCCCAAT 699
700 CATGAGCCATGCCCCCGGCGCTGCTTACAGAGGAGCTTTGCGCCCTGCAAAAGG 759
372 CATGAGCCATGCCCCCGGCGCTGCTTACAGAGGAGCTTTGCGCCCTGCAAAAGG 759
760 CCAGATCGGATCTCGCTCAACGCGGAGCTTATGAGCCCTGCGGACAGCAATGAGCTCG 819
432 CCAGATCGGATCTCGCTCAACGCGGAGCTTATGAGCCCTGCGGACAGCAATGAGCTCG 819
820 GGAAGAGAGGCTGAGCGAGCGATGGAATTTCACTTGGCTGTTGGCCATCCCAT 879
492 GGAAGAGAGGCTGAGCGAGCGATGGAATTTCACTTGGCTGTTGGCCATCCCAT 879
880 CTTCTTGAAGAAGACATCAAGAGAGCAAGAAAGAGAGCTGGGCGAGAGGCTTCCAGC 939
552 CTTCTTGAAGAAGACATCAAGAGAGCAAGAAAGAGAGCTGGGCGAGAGGCTTCCAGC 939
940 CTTCACTCCGCGGAGCTTTGCGCATCTCTCAATGCGGAGAGACCGACTTCTACGCAATGA 999
612 CTTCACTCCGCGGAGCTTTGCGCATCTCTCAATGCGGAGAGACCGACTTCTACGCAATGA 999
1000 TTAATAACATCCCAATTTCCGCGGCGCACTTGAAGCGGTCCCGTCCGAGAGCGATATCT 1059
672 TTAATAACATCCCAATTTCCGCGGCGCACTTGAAGCGGTCCCGTCCGAGAGCGATATCT 1059
1060 CCGGCGCATCATGAGACCGAGAGATTAAGAGCGGAGGCC 1101
732 CCGGCGCATCATGAGACCGAGAGATTAAGAGCGGAGGCC 1101

TITLE
JOURNAL
COMMENT
FEATURES
source
1..703
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="Qme6"
/db_xref="taxon:51453"
/clone="tric039xe13"
/dev_stage="mycelia"
/clone_lib="T. reesei mycelial culture, Version 6 October 2003"
/note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN
Query Match 35.2%; Score 701; DB 14; Length 703;
Best Local Similarity 99.7%; Pred. No. 8.1e-114; Indels 0; Gaps 0;
Matches 701; Conservative 1; Mismatches 1;

399 GCGGGGATACAGCCTTGGGTGATCTTTGATCACTGGGATCTGCTCAGGCGCTTACG 458
1 GCGGGGATACAGCCTTGGGTGATCTTTGATCACTGGGATCTGCTCAGGCGCTTACG 458
459 ATCGCTATGAGGCTGCTCAACGATGAGAGGTCAGCTGAGCTTTGAGCGGTATGCGA 518
61 ATCGCTATGAGGCTGCTCAACGATGAGAGGTCAGCTGAGCTTTGAGCGGTATGCGA 518
519 GATTGTGCTTTGAAAGTTTGGGAGCGAGTCCAGAACTGGATCAACATCAACGAMCCCT 578
121 GATTGTGCTTTGAAAGTTTGGGAGCGAGTCCAGAACTGGATCAACATCAACGAMCCCT 578
579 GATTGTGCTTTGAAAGTTTGGGAGCGAGTCCAGAACTGGATCAACATCAACGAMCCCT 578
181 GATTGTGCTTTGAAAGTTTGGGAGCGAGTCCAGAACTGGATCAACATCAACGAMCCCT 578
639 ACAAGCACTCCACCGAGGAGGCAACCTGCACTGAGCGGTGCTGCTGGAAGGCCCA 698
241 ACAAGCACTCCACCGAGGAGGCAACCTGCACTGAGCGGTGCTGCTGGAAGGCCCA 698
699 TCATGAGCCATGCGCGCGCGTGGCGCTTACAGAGGAGCTTTGCGCCCTGCAAAAGG 758
301 TCATGAGCCATGCGCGCGCGTGGCGCTTACAGAGGAGCTTTGCGCCCTGCAAAAGG 758
759 GCCAGATGCGCATCTGCTCAACGCGGAGCTTACTATGAGCCCTGGGAGCAAGTGAAGCTC 818
361 GCCAGATGCGCATCTGCTCAACGCGGAGCTTACTATGAGCCCTGGGAGCAAGTGAAGCTC 818

QY 819 GGGACAAAGAGGCTGCTGAGCCGAGATGGAATTTCAATTGGCTGTTGGCAATCCCA 878
 DB 421 GGGACAAAGAGGCTGCTGAGCCGAGATGGAATTTCAATTGGCTGTTGGCAATCCCA 480
 QY 879 TCTTTTGAAGAAGACTATCCAGAGAGCATGAAGAAGAGCTGGCGAGAGGCTTCCAG 938
 DB 481 TCTTTTGAAGAAGACTATCCAGAGAGCATGAAGAAGAGCTGGCGAGAGGCTTCCAG 540
 QY 939 CCTCTACTCCCGGAGCTTTGGCCATCTCAATCCCGAGAGACCGACTTCTACGCGATGA 998
 DB 541 CCTCTACTCCCGGAGCTTTGGCCATCTCAATCCCGAGAGACCGACTTCTACGCGATGA 600
 QY 999 ATTACTACACATCCCATCTGCGCGCCACTTACGAGTCCCGCCGAGAGGAGCATATC 1058
 DB 601 ATTACTACACATCTCCAGTTGCGCGCCACTTACGAGTCCCGCCGAGAGGAGCATATC 660
 QY 1059 TCGGCGCCATCCATGAGCAGCAGAGAAATMAAGACGCGACGCC 1101
 DB 661 TCGGCGCCATCCATGAGCAGCAGAGAAATMAAGACGCGACGCC 703
 RESULT 3
 CF884351 584 bp mRNA linear EST 31-OCT-2003
 LOCUS tric039xel3.b12 T. reesei mycelial culture, Version 6 October 2003
 DEFINITION Hypocrea jecorina cDNA clone tric039xel3, mRNA sequence.
 ACCESSION CF884351
 VERSION CF884351.1 GI:38139033
 KEYWORDS EST.
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
 ORGANISM Hypocrea jecorina
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
 1 (bases 1 to 584)
 Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houlek, T.D.,
 Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
 Dean, R.A.
 Analysis of the protein processing and secretion pathways in a
 Trichoderma reesei EST database
 Trichoderma reesei
 Contact: Ralph A. Dean
 Fungal Genomics Laboratory
 North Carolina State University
 Campus Box 7251, Raleigh, NC 27695, USA
 Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph.dean@ncsu.edu
 Seg primer: Tr-F1 primer
 Location/Qualifiers
 1..584
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QMGa"
 /db_xref="taxon:51453"
 /clone="tric039xel3"
 /dev_stage="mycelia"
 /clone_lib="T.reesei mycelial culture, Version 6 October 2003"
 /note="Vector: PREP; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 23.2%; Score 461.2; DB 14; Length 584;
 Best local Similarity 94.0%; Pred. No. 2e-71;
 Matches 489; Conservative 1; Mismatches 29; Indels 1; Gaps 1;

QY 399 GGGGGGATACGCTTGGGTGATTTGACACATGGGATCTGCTCAGGCGCTTACAG 458
 DB 1 GGGGGGATACGCTTGGGTGATTTGACACATGGGATCTGCTCAGGCGCTTACAG 60
 QY 459 ATGCGTATGAGAGCTGCTCAACGTGGAAGAGTCCAGCTTGACCTTGACGCGATGCGA 518

DB 61 ATGCGTATGAGAGCTGCTCAACGTGGAAGAGTCCAGCTTGACCTTGACGCGTATGCGA 120
 QY 519 GGTGTGCTTTGAACGCTTTTGGGAGACCGAGTCCAGAACTGATACATCAACAGACCTT 578
 DB 121 GGTGTGCTTTGAACGCTTTTGGGAGACCGAGTCCAGAACTGATACATCAACAGACCTT 180
 QY 579 GGAATTAGGCGCATCTATGATATATGCCACCGGAGCAGAACCGCCCGGAGAGGAGCATTA 638
 DB 181 GCATTAGGCGCATCTATGATATATGCCACCGGAGCAGAACCGCCCGGAGAGGAGCATTA 240
 QY 639 ACAAGCACTCCACCGAGGCAACACTGCACTGAGCCGTGCTCGTGAAGAGGCCAGA 698
 DB 241 ACAAGCACTCCACCGAGGCAACACTGCACTGAGCCGTGCTCGTGAAGAGGCCAGA 300
 QY 699 TCATGAGCCATGCGCGCGCGCGCGCTGCTTACAGCAGGAGACTTCCGCCCTCCAGAAAG 758
 DB 301 TCATGAGCCATGCGCGCGCGCGCGCTGCTTACAGCAGGAGACTTCCGCCCTCCAGACGG 360
 QY 759 GCGAGATCGGCATCTGCTCAACGCGGACCTATGAGCCCTGGAGCAGCATGAGCCTC 818
 DB 361 GCGAGATCGGCATCTGCTCAACGCGGACCTATGAGCCCTGGAGCAGCATGAGCCTC 420
 QY 819 GGGACAAAGAGGCTGCTGAGCCGAGATGGAATTTCAATTGGCTGTTGG-CCAAATCCC 877
 DB 421 GGGACAAAGAGGCTGCTGAGCCGAGATGGAATTTCAATTGGCTGTTGGCCCAATTCCC 480
 QY 878 ATTTTCTTGAAGAAGACTATCCAGAGAGCATGAAGAAGC 917
 DB 481 ATCTCTTGACACGAGCATATCCAGACAGCATGCAATCAAC 520

RESULT 4
 AA415086 1994 bp mRNA linear EST 09-DEC-1999
 LOCUS Mg0020 RCW lambda Zap Express Library Magnaporthe grisea cDNA clone
 DEFINITION RCW20 similar to Beta-Glucosidase (EC 3.2.1.21), mRNA sequence.
 ACCESSION AA415086
 VERSION AA415086.1 GI:2537251
 KEYWORDS EST.
 SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)
 ORGANISM Magnaporthe grisea
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetidae; Incertae sedis; Magnaportheaceae; Magnaporthe.
 1 (bases 1 to 1994)
 Wu, S.-C., Bernstein, B.D., Dervill, A.G. and Albersheim, P.
 Expressed sequence tags of the rice blast fungus grown on rice cell
 walls
 Unpublished (1997)
 Contact: Sheng-Cheng Wu
 CCRRC
 University of Georgia
 220 Riverbend Road, Athens, GA 30602-4712, USA
 Tel: 706 542 4446
 Fax: 706 542 4412
 Email: wus@ccr.ucr.edu
 Fully sequenced
 Insert Length: 1994 Std Error: 0.00.
 Location/Qualifiers
 1..1994
 /organism="Magnaporthe grisea"
 /mol_type="mRNA"
 /strain="CP987"
 /db_xref="taxon:148305"
 /clone="RCW20"
 /sex="Mati-1 hermaphrodite"
 /tissue_type="Mycelium"
 /dev_stage="Day 5 post-inoculation"
 /clone_lib="RCW lambda Zap Express Library"
 /note="Vector: pBluescript excised from lambda Zap
 Express; Site 1: EcoRI; Site 2: XhoI; Day 5
 post-inoculation mRNA prepared from Magnaporthe grisea
 grown at 23C in the dark with constant gyratory shaking

FEATURES

source
 1..1994
 /organism="Magnaporthe grisea"
 /mol_type="mRNA"
 /strain="CP987"
 /db_xref="taxon:148305"
 /clone="RCW20"
 /sex="Mati-1 hermaphrodite"
 /tissue_type="Mycelium"
 /dev_stage="Day 5 post-inoculation"
 /clone_lib="RCW lambda Zap Express Library"
 /note="Vector: pBluescript excised from lambda Zap
 Express; Site 1: EcoRI; Site 2: XhoI; Day 5
 post-inoculation mRNA prepared from Magnaporthe grisea
 grown at 23C in the dark with constant gyratory shaking

100 rpm in Vogel's minimal medium containing 0.5% isolated rice cell walls as the sole carbon source. Library provided by Sheng-Cheng Wu. Sequences were processed by one of two methods. Where a full-length alignment to the M. grisea genome sequence was available, the EST sequence was trimmed according to the alignment, otherwise sequence quality was assessed using phredPhrap version 991019 and trimmed according to phd files (0.05) and for vector seqs."

ORIGIN

Query Match 19.4%; Score 387; DB 9; Length 1994;
Best Local Similarity 57.0%; Pred. No. 1.8e-58;
Matches 806; Conservative 1; Mismatches 591; Indels 15; Gaps 5;

```

Qy 78 CTCTGCCCAACGACTTTGAAATGGGGCTTCGACAGCGCCCTTACACATGGAAGCCCG 137
Db 290 CTCTTCCCAAGATTTCTTTGGGAGCTTCGACAGCGCTCATACAGATCCAGAGGTGCTA 349
Qy 138 TCAAGAGAGTGGCCGCGCCGCTTCATCTGGGACAGCTACCTGAGGCGCATGCG 197
Db 350 TCGACAGAGATGCTGCTGGCCGCTTCATCTGGGATACCTTACGCAATTCGGGTAAAG 409
Qy 198 GCACCAACGCGCCCAAGCGCATGTGCTTGCATCACTACACCGCTTACGATGAGACT 257
Db 410 TCGCAGACGCGCAGCTCGGAGTGTGACGGCATGCACTACACAGAGACGACGAGAGCA 469
Qy 258 TTGATCTTTGACCAAGTACGAGCGCAAGGCGCTACCGCTTCTTGCGGTGCTGCGGA 317
Db 470 TTGACCTGCTCAAGTCTGTTGGTCCCAAGTACGTAATGCTTTTCATCTCGTGTGAGGA 529
Qy 318 TCATTCCTCTCGCGCGGAGCGCTGATCCGCTCAACGAGAGGAATGATTTACAGCA 377
Db 530 TTATCCCATATGGTGGCCGTATATAGCCCATCAACGAGAGGATGACCATCACTGCA 589
Qy 378 AACTGATTTACGCTCTGTTGAGCGGGATTCACGCTTGGGTACTTTGTACACTGGG 437
Db 590 AGTTCGTAGACGACTGCTCGAGGCGCGCATCAAGCCATCATACCTCTTCCACTGGG 649
Qy 438 ATTCGCTCCAGGGGCTTCAGATGCTATGAGAGCTGGCTCAAGTGAANAAGTCCAGC 497
Db 650 ATTCGCGGAGTATGAGACAGCGGTACGCTGGGTGCTGAAACAGGAAAGTTCCTCT 709
Qy 498 TGAATCTTGAAGCGGTATGCGAGGTGTGCTTGAACGTTTGGGAGCAGATCCAGAACT 557
Db 710 TGAATCTTGAAGCACTATGACAGCGCTCATGTTCAAGGGCAATCCAGAG---CAAACCT 766
Qy 558 GGATCACCATCAAGAMCCTTGATTCAGGCCATCTATGATATGCCACCGGCAAGCAAG 617
Db 767 GGATCACCCTTCAGAGCGGTGTGCTGCTCATTTCTGCTTACAGGGTGGTCAAGTTTG 826
Qy 618 CCGCGGAGGAGAGAGCATTAACAGCACTCCAGCGGCAACAGTCCGCACTGAGCGCT 677
Db 827 CGCTTGGCGCTGTCTCGACCGCAGAGTGGCTGTGCGGAGACAGCTTCGGGAGCGCT 886
Qy 678 GGCCTGCTGGAAGAGCCAGATCATGAGCCATGCCCCGCGCGCTGTCAACAGCAGAG 737
Db 887 GGATTTGGGCGCACACTTCTGTTGGCCACGCTCGTGGCTCAAGGTGACAGGGAAG 946
Qy 738 ACTTTGCGCCCTTCGCAAAAGGCGCAGATCGGCATCTGCTCAACGCGCACTATGAGC 797
Db 947 AATTCAAGCGCAGACAGAGGTGATGATGATCAAGTTGAAAGGAGGAGCCCACTTCC 1006
Qy 798 CCGTGGACAGCAATGAGGCTCGGACAAAGAGGTGCTGAGCGACGAGTGAATTTGACA 857
Db 1007 CTTGGGATCCCGAAGACCCCAAGGAGCTGCTGATGACGACCAACGCAAGTTCAGTTCCGA 1066
Qy 858 TTGGCTGTTTGCATTCCTCTTCTTGAAGAAGACTATCAAGAGCATGAGAGAGC 917
Db 1067 TCTGTGTTTGGGACCCCATCTACTTTGGCGAG---TATCCAGTCTCGATGGTAAAGC 1123
Qy 918 AGCTGGCGAGAGGCTTCAGCCCTCACTCCCGGAGCTTTGCAATCTCAATCCGAGAG 977

```

```

Db 1124 AGCTTGGCGACCGCTGCCCCACTTTTCAACGAGAGAGAGGAAGGCTTTAGTCA---GGGTT 1180
Qy 978 AGACCGACTTCTACGGCATGAATTTACTACACATCCGATTCGGGCGGCACTTGAAGGCTC 1037
Db 1181 CAACGACCTTTATGGAATAACTGCTACCGGCACTCACTACAGCACAAGAGAGAG 1240
Qy 1038 CCGTCCCGGAGAGGAGACTATCTCGGCGCCATTCATGACACAGAGAAATGAAGCGCA 1097
Db 1241 AGCGGCGCGAGAGAGATTAACCTTTGAAACCTGGAAAGTGTCTTCAACAAGAGCGGCG 1300
Qy 1098 GCGCCGCTGGCGAGAGAGAGGCGCTGCGCTGCTGCGCTCTCTGCGGACATGTCGGA 1157
Db 1301 AATGCACTGCGCCCGGAGACACAGTCAACCGGCTTCTTCAACAGCAAGAGGTTCGAG 1360
Qy 1158 AGCATCTGCGCGGGGTATAGGCTGTAAGGCAAGCCC---ATTCATCATCCAGAAAGC 1214
Db 1361 AGCTGCTGTTTGGCTCAGCAAGCGTTTCACTACCCAGATCTTGGTTACGAGAAATG 1420
Qy 1215 GATGCCCGTCCCTGAGAGAGAAACATGACGTGCGAGAGAGGCGCTCAACGACCCCTTC 1274
Db 1421 GCAAGTGTGTCMAAGGGGAGAGAGATGCTGCGCTCGAAGAGATCTGAGAGATGCTTC 1480
Qy 1275 GCATCGSTACTTTGACTCGCACTGGAATCTGCAATTTCCAAAGGCACTTACCAGAGCGG 1334
Db 1481 GGGTCCAGTACTACAGATATATGTCAAAGGCTTTGGCGCGGCTTATTCGAGAGATGTC 1540
Qy 1335 TCGTGTCAAGGGGTACTTTGCGGGCGCTTGTGCTGATTAATCTTGAATGATGATGCT 1394
Db 1541 TAAACGTCCGGGATCTCAAGCTTGGTGTGATGACAACTTTGATGGGCGGAGGCT 1600
Qy 1395 ACAGACCCAGATTCGCGGTACGTTCAACAGACTAC---ACCACCTTCAAGCGGCA 1451
Db 1601 ACAGACCAAGATTTGGTGTGACCTTTGTGACTACGAGAAACGCGCAGAAAGCTTACCCA 1660
Qy 1452 AGAAGTGTGCTGCTGCTTCAAGACATGTTTG 1484
Db 1661 AGAAGAGCGCCAAAGCAATGAAGCCGTTGTTTG 1693

```

RESULT 5

AY109400 2031 bp mRNA 1linear HTC 17-OCT-2002

LOCUS AY109400 Zea mays CL1052_1 mRNA sequence.

DEFINITION Zea mays CL1052_1 mRNA sequence.

ACCESSION AY109400

VERSION AY109400.1 GI:21213111

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. .2031

/organism="Zea mays"

/mol_type="mRNA"

location/Qualifiers


```

/db_xref="taxon:39947"
/clone="OSJNBe13N03"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNBe"
/notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

```

ORIGIN

```

Query Match      9.8%; Score 195.2; DB 14; Length 786;
Best Local Similarity 56.3%; Pred. No. 1.8e-24;
Matches 409; Conservative 0; Mismatches 308; Indels 9; Gaps 2;

OY 81 TCCCCACGACTTTAATGAGGCTTCGCAACGGCGCCCTACGAGATCGAAGGCGCCGCA 140
DB 60 TCCCGGAGAGCTTCAATCTTCGCGACCGGCTCGCGCTTATAGTACGAAGCGCTGTGA 119
OY 141 AAGAAAGTGGCCGCGGCGCCCTCATCTGGGACACGTAAGCTGCAAGCCATGCGCA 200
DB 120 ACGAAGCGCGGACGAGCGCGGACATCTGGGACACGTAAGCTGCAAGCCATGCGCA 179
OY 201 CCAACGGCGGCAACGCGGCAATGCGCTTGATCAACCGCTACGATGAGACTTGG 260
DB 180 AAGATGTAAGCAAGGAGATGAGCAATGATTTTACATGCTACAGAGAGATCTGA 239
OY 261 ATCTCTTGAACCAAGTACGAGCGCGCAAGGCTTACCGCTTCTCTGTCGTGCGGATCA 320
DB 240 ACTTCGTACCGACATGAAACATGACGCGCTTCCTTCATTCATTCCTTGGAGAGAGATCC 299
OY 321 TTTCCCTCGGCGGCGAGCTGATCCCGTCAACGAGAGAGGAATGATTTTACAGCAAAAC 380
DB 300 TCCCAAAATGAAACCATCAGTGAAGGAATCAACAAAGAGGATGCTTTTCAACAAAGCC 359
OY 381 TGATGACGCGCTGTTGAGCGGCGGATATCAAGCTTGGGTGATCTTGATCAACGAGGATC 440
DB 360 TGATTAAGAGATCATATCTAGAGGCTTGAAGCCATTTGTCAATCTTCAATTTGACA 419
OY 441 TGCTCAGAGCGCTTACGATGCTATGAGAGCGCTGCTCAACGTAAGAGGTCAGCTGG 500
DB 420 CCCCACAGGCTCTGGAGAGCAAAATACCGAGCTTCTCA---GTGAAAACATTTGAAAG 476
OY 501 ACTTGAAGCGGATGAGAGGTTGCTTGAACGTTTGGGAGACGAGTCCAGAACTGGA 560
DB 477 ATTTGCTGAGACTATGCGAGCGTGTCTTCGTGAGTTGGGAGACCGGATGAAGTCATGGA 536
OY 561 TCACCATCAACGAMCCCTGATTCAGGCGCATCTATGATATGCAACCGGAGCAACGCGCC 620
DB 537 ACAAGTTCAAGAGCGCATGATCTTCTGCGCGCGGCTTACGCTTCGCGACCAAGGCCC 596
OY 621 CGGCGAGAGCAG-----CATTAACAAGCACTCCACGAGGAGCAACCTGCACTGAGC 674
DB 597 CCGGCGCGCTGCTCGCGCTGACGCTTCCAAAGAAATCGCGCCCGGAGCACTCCGCAACGAG 656
OY 675 CGTGCGCTGCTGGAAGGCGCAGATCAAGGCAATGAGCCGCGCGCGCTGCGCTTACAGCA 734
DB 657 CGTACGTGCGCGGCAACAACCTGCTGCTCGCCAGCGCGAGGCGCTGCGCTTACAGCGCC 716
OY 735 GGAACCTTTGCGCCCTCGCAAAAGGCGCAGATCGGCAATCTGCAACGAGCGACTATATG 794
DB 717 AAAAGTACGAGCGGACGAGAAAGGCGGAGATCGGCAATCAAGCAGGTGTGCACTGTTGG 776
OY 795 AGCCCT 800
DB 777 TGCCCT 782

```

```

RESULT 7
LOCUS CB673495 786 bp mRNA linear EST 09-Apr-2003
DEFINITION OSJNBe08A17.f OSJNBe Oryza sativa (japonica cultivar-group) cDNA
ACCESSION CB673495

```

```

VERSION CB673495.1 GI:29677220
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretidae; Oryzae; Oryza.
1 (bases 1 to 786)
REFERENCE Jantaauriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Ming,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
COMMENT Contact: Rod Ming
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 08 row: A column: 17
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1..786
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNBe08A17"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNBe"
/notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

```

FEATURES

```

source
1..786
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNBe08A17"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNBe"
/notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

```

ORIGIN

```

Query Match      9.7%; Score 193; DB 14; Length 786;
Best Local Similarity 57.1%; Pred. No. 4.4e-24;
Matches 437; Conservative 0; Mismatches 310; Indels 18; Gaps 4;

OY 127 CGAAGCGCGCTGCAAGAGGTGCGCGCGCGCTGATCCGTCATCTGGGAGACGTAAGCCACT 186
DB 22 CAGGAGCTGTGAAGAGAGCGGAGAGGCGACATCTGGGACACGTTGCGGCAAC 81
OY 187 GAGCCATGCGGACCAAGCGCGCAACGCGATGTCGCTTGCATCACTACCAACGCTA 246
DB 82 CTTGGAAGATCACCGACTTCAAGCAATGTCATGTCAGTTGATCAATCAACCGTTT 141
OY 247 CGATGAGACTTGAATCTTGAACCAAGTACGCGCGCAAGGCGCTTACCGCTTCTTCTTGC 306
DB 142 CGAGAGGATATACACTATGACATGAGGATGATGATGATGATGATGATGATGATGATGATG 201
OY 307 GTGTCGCGGATCAATCCCTCGCGCGGAGGCTGATCCGTCACAGAGAGGAAATTA 366
DB 202 ATGCTCAAGATTTACCAATGATG-----TTGCTCAAGTCAATCAAGCTGATGCA 255
OY 367 GTTTACAGCAACATGATGACGCGCTGTTGAGCGCGGATCAAGCTTGGGTGACTTT 426
DB 256 CCACTACAAACAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 315
OY 427 GTACCACTGAGATCTGCTTCAAGCGCTTACAGATGCTATGAGAGTGTGCTCAAGTGA 486
DB 316 CTACCACTGAGACTTCCCGAGGCGCTTGAAGCAAGTCAAGGCGTGTGAGAGG-- 373
OY 487 AGAGGTCCAGCTGACTTGAAGCGGATGCGAGGTTGCTTGAACGTTTGGGAGCG 546
DB 374 -CAGATAGTGAAGATTTGGCGCGGCTGACCGGAGAGCTGCTTCAAGGAGTTGGGAGCAG 432

```


QY 547 AGTCCAGAACTGATCCATCAACGAMCCCTGATTACAGCCATCTATGATATGCCAC 606
DB 433 GGTGAAGACATGATGATCAAGCTCAACGAGCGCACACGCTGCGATCCAGGGCTACGACGC 492
QY 607 CGGCAGCAAGCCCCCGGGGAGAGCAGATTA-----ACAAGACATCTCCACGAGGGCAA 660
DB 493 AGGAGCTCAAGCCCCCGGGGAGAGCAGATTA-----ACAAGACATCTCCACGAGGGCAA 552
QY 661 CACTGCACATGAGACCCGCTGCTGCTGGAAGGCCACAGATCATGAGCCATGCCCCGCGT 720
DB 553 CTCGGGACCGAGACCTTACGTCCTGCTGCTGGAAGGCCACAGATCATGAGCCATGCCCCGCGT 612
QY 721 GCGCGTCTACAGAGGAGACTTTGCGCCCTGCAAAAGAGGCGAGATCGGATCTGCTCAA 780
DB 613 CACATCTACAGAGCAAAATACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 672
QY 781 CGCGGACTACTATGAGACCCCTGGAACAGATGAGACCTCGGAGCAAGAGGCTGTGAGCG 840
DB 673 CGTGAATGTGATGAGAGCCGATGTCACACA---CCACATGACATGAGAGGCGGCAAGAG 729
QY 841 ACGGATGGAATTTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 885
DB 730 AGCGCAGAGATTTACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 774

RESULT 8 CB642881 806 bp mRNA linear EST 08-Apr-2003
LOCUS OSJNB03E17.f OSJNB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION Oryza sativa (japonica cultivar-group) cDNA
ACCESSION CB642881
VERSION CB642881.1 GI:29637872
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriarctideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 806)
Jantschuliyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
JOURNAL Contact: Rod Wing
COMMENT Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tac gac cat g
Plate: 03 row: B column: 17
Seq primer: gta aac cga cgg cca gtc.
FEATURES
source
1..806
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultiar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB03E17"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_1ib="OSJNB"
/note="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2:
XhoI, 24 hrs after inoculation with Rice Blast (Che
86061)"

Query Match 9.6%; Score 190.4; DB 14; Length 806;
Best Local Similarity 57.5%; Pred. No. 1,3e-23;
Matches 407; Conservative 0; Mismatches 286; Indels 15; Gaps 3;

QY 81 TGCCCAACGACTTTGATGAGGCTTTGCAACGCGCCCTTACCGAGATCGAAGGCGCGCTCA 140
DB 96 TCCCGAGGGGCTGCTTTCGAGCAGCGCTTCCCGCCGATACAGATCGAAGGAGCTGTGA 155
QY 141 AAGAGGTGGCGCGGCGCGCTTCCATCTGGGACAGCTACTCCACCTGAGGCCATTCGCGCA 200
DB 156 AGAGAGACGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 215
QY 201 CCAAGCGCGCAACGAGAGTGTGCTTGGATCACTACCAACCGCTAGATGAGAGACTTTCG 260
DB 216 CCGACTTACAGCAATGCTGATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 275
QY 261 ATCTCTTACCAAGTACGCGGCAAGGCTTACCGCTTCTCTTGTGTGTGCTGCGGATCA 320
DB 276 AGCTCATGCGAGACATGAGGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 335
QY 321 TTCCTCGGCGGAGAGGCTGATCCGCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 380
DB 336 ACCCAATAGGTG-----TTGTCAGATCATACAGCTGTGTATGACACTTACCAACAGC 389
QY 381 TGAATGACGCTCTTGTGAAGCGGGATATACAGCTTGTGATCTTGTATCACTGAGATC 440
DB 390 TGATGATGATGATCTTACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 449
QY 441 TGCTTACGCGCTTTCACGATGCTATGAGAGCTGTGCTCAAGCTGAGAGAGGCTCAGCTGG 500
DB 450 TCCCGAGGCGCTTGAAGACAGATACAGAGGCTGTGACAGG---CAATATGTGAGC 506
QY 501 ACTTGAAGCGGATGAGAGGTTGTGCTTGAACGTTTGGGAGCCGAGTCCAGAACCTGGA 560
DB 507 ATTTCGCGGCTACGCGAGACGCTGCTTACGAGAGTTCCGAGAGGAGGAGGAGGAGGAGG 566
QY 561 TCACATCAACGAGACCTGATTCAGGCACTATGATGATGATGATGATGATGATGATGATGAT 620
DB 567 TCACGCTCAACGAGCGGACACAGGTGGCATCCAGGGCTACGAGCGAGGCTCAAGGCC 626
QY 621 CGGCGAGAGGAGC-----ATTACAGACCTTCCACCGAGGCAACACTGCACTGAGC 674
DB 627 CCGCGCGCTGCTTCCGCTGCTTCCACCTTACCTGACGAGGCGGCAACTCCGCGACCGAGC 686
QY 675 CGTGGCTCGCTGAAGGCCAGATCATGAGCCATCCCGCGCGCGCTTACAGCA 734
DB 687 CTAAGCTGCTGCGCCACACTTCTGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746
QY 735 GCGACTTTGCGCCCTGCAAAAGGCGCAGATCGGATCTGCTCAAGC 782
DB 747 CAATATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 794

RESULT 9 CD438244 889 bp mRNA linear EST 03-JUN-2003
LOCUS EL01N0511A05.b EndospERM_5 Zea mays cDNA, mRNA sequence.
DEFINITION CD438244
ACCESSION CD438244
VERSION CD438244.1 GI:31353887
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoidae; Andropogonae; Zea.
REFERENCE 1 (bases 1 to 889)
Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
Messing, J.
Sequencing of the maize endospERM ESTs
Unpublished (2002)
JOURNAL Contact: Lai, Jinseng
COMMENT Dr. Joachim Messing's lab
Wakeman Institute, Rutgers University

190 Prellinghuysen Rd., Piscataway, NJ 08854, USA

Tel: 732-445-3801

Fax: 732-445-5735

Email: j1a1@wakeman.rutgers.edu

Seq primer: T3.

FEATURES

source

Location/Qualifiers
1. 889
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm 5"
/note="Vector: pbluescript SK-, site_1: EcoRI; site_2: XhoI"

ORIGIN

Query Match 9.0%; Score 179.2; DB 14; Length 889;
Best Local Similarity 54.8%; Pred. No. 1.2e-21;
Matches 493; Conservative 0; Mismatches 383; Indels 24; Gaps 6;

```

QY 113 GCCGCTACACAGATCGAAGGCGCGCTCAAGAGGTGCGCGCCGTCATCTGGGAC 172
Db 5 GCGGTGATCCAGATCGAGGCGCGCGCTCAAGAGGTGCGCGCCGTCATCTGGGAT 64
QY 173 ACCTACTGCACTGAGGACATCGGCAACCAAGCGGCAACGCGATGCGCTTGGCAT 232
Db 65 TCATTCGCGCAAGTCCAGAAATATTCAGAGAAATCAAAATGAGAGCTTGCACTGAT 124
QY 233 CACTACCAACCGCTACGATGAGGACTTGTATCTCTTGAACCAAGTACGCGCAAGGCTTAC 292
Db 125 CAATTCATCCGCTACAGAAAGACCTCGATCTCAAGAAAGTTGAATCTTGAATGCTTAC 184
QY 293 CGCTTCCTGCTGCTGCTGCGGATCATTCCTCGCGGCGGAGCTGATCCGTCAC 352
Db 185 CGGTCTCATCTCATGTGCTCCAGAGATCTCCGAGTGGCGAAGG-----AAATCAT 238
QY 353 GAGAGGGAATGAGTTTACGCAACATGATGACGCGCTTGAGGCGGGGTATCAG 412
Db 239 CCAGAGGTGATGCTATTAACAATATTTGATTAATCTTCTGCTTCAAGAGGATGACT 298
QY 413 CTTTGGGTGACTTTGTACCACTGGGATGCTGCTCAGGCGCTTACGATCGCTATGAGGC 472
Db 299 CCTTACATCAACCTTACCACTATGATCTTCTCTTGCGCTTGAGAAATATGAGAGG 358
QY 473 TGGCTCAACGTGGAAGAGTCCAGCTGACCTTGGACCGGTATGCGAGTTGCTTTGAA 532
Db 359 TGGTTAAGCGCGAAGATGCGGAGCTTG-----TTTACAGACTATGACTTCTGTTTAA 415
QY 533 CGTTTGGGGAACGAGTCCAGAACTGATCAACATCAAGCAAGCCCTGATTCAGGCCATC 592
Db 416 ACCTACGGGATCGGTAAGACATGCTTTTACATTCATGAGCCAGAGATGATGAGCTTA 475
QY 593 TATGATATGCAACCGGAGCAACGCCCGGCGAGAGAGCATTTAACAGACCTCCACC 652
Db 476 CTGGCTATGACACAGGATCAATCTCTCAAAAGTGC-----ACCAGATGCGCTGCT 529
QY 653 GAGGCAACACTGCACTGAGCGGCTGCTGCTGGAAGGCCAATATGAGCCATGCC 712
Db 530 GGTGGGAATTCAGCAACCGAATTTAACAATGTTGCTCAATATTTTCTTGGCAATGT 589
QY 713 GCGGCGTGGCGCTGACAGAGGAGCTTGGCCCTCGCAAAAGGCGCAGATCGGCAATC 772
Db 590 ACTGCAATTTGCAAGATACCTGACGAATATCAGGCTGCTCAGAAAGGTATGATCGAATA 649
QY 773 TCGCTCAACGCGACTACTATGAGCCCTGGAACAGCAATGAGCTTCGGAGCAAGAGCT 832
Db 650 GTCCTGGAATTCACCTGATGAGGCTTT--ACAACTCACTGATGATGACCAAGAGCA 706
QY 833 GCTGAGCAGCGATGGAATTTGACATTTGGCTGCTTTGCAATCCATCTTCTTGAAGAG 892
Db 707 GCCCAAGAGCGAGGACTTCAATTTGCTGCTGTTTGTATTC--ATTGATTAACGGA 763

```

```

QY 893 GACTATCCAGAGAGCATGAGAGAGCAGCTGGCGAGAGGCTTCCAGCCCTCACTCCGCG 952
Db 764 CACTATCCAGAGATGATGAGATCTCGTGAAGAGAGGCTGCCAGGTTCACTCTCGAG 823
QY 953 GACTTGGCACTCTCAATGCGGAGAGACCGACTTCTACGCGCATGATTTCTACATCC 1012
Db 824 CA---GGCTAACTGATGAGGCTGCGGAGCATCAATCGGTATCAACGATCAACATCC 880

```

RESULT 10

CB688082

LOCUS

DEFINITION

CB688082 480 bp mRNA linear EST 04-SEP-2003
CBST-23-E-06 Mixed source, strain EP15 and EP15 infected with
hypovirus CHV1-EP713 Cryphonectria parasitica cDNA clone EP15,
EP15-CHV1-EP713 5-prime, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CONTACT

UNIVERSITY

TELEPHONE

FAX

EMAIL

LOCATION

REMARKS

NOTES

DESCRIPTION

FEATURES

source

1. 480
/organism="Cryphonectria parasitica"
/mol_type="mRNA"
/strain="EP15"
/db_xref="ATCC (insect):38755 and 52571"
/db_xref="taxon:5116"
/clone="EP15", EP15-CHV1-EP713"
/clone_lib="Mixed source, strain EP15 and EP15 infected
with hypovirus CHV1-EP713"
/note="Cryphonectria parasitica EST derived from mixed
samples of cDNA from wild-type (strain EP15) and
hypovirus-infected cultures"

ORIGIN

Query Match 8.8%; Score 175.6; DB 14; Length 480;
Best Local Similarity 64.4%; Pred. No. 6e-21;
Matches 290; Conservative 0; Mismatches 158; Indels 2; Gaps 2;

```

QY 72 CGTAGCTCTGCCAAGCACTTTGATGAGGCTTGCACGCGCGCTTACCAATCGAAG 131
Db 4 CGGTCGCTACATCAATCAAGTGGGCTTTGCGACGCGCTCTTACCAAGGTGCAAG 63
QY 132 GCGGCTTAAAGAGTGGCGCGCGCGCTTCACTGGAACGATCGCAACCTGAGAC 191
Db 64 GCTTACCGAAGAGAGCGCGCGCGCTTCACTGGAACGATCGCAACCTGAGAC 123
QY 192 CATGCGCAACCAAGCGCGCGCGCTTGTGCTTGGATCACTACCAACCGCTTACGATG 251
Db 124 GCAAGATGCGGAGCGGCTTCAAGCGGCGGCTGCTGGAAGATCAACCGCAAGCGGCG 183

```

OY	252	AGGATTTGATCTTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTTGATGAGT	311
Db	184	AGGATTTGCGCTGCTCAAGAGCTCGGGGCGACGGCGTACCGGTTCTCGCTGTGAGT	243
OY	312	CGCGGATATTTCCCTCGGCGGACGGCTGATCCCGTCACAGAGAGGAAATTGATTTT	371
Db	244	CGCGGATATCTCCGCTGGCGGGCGCGACGACCCCGGTCAACCAAAAAGGGGCTGGACCACT	303
OY	372	ACAGCAACTGATTTGACGCGCTGTTGAGGCGGGGATATCAGCGCTTGGGTATTTGTATCC	431
Db	304	ACAGCGCGTTACCAACAGACTCTGCTGGCGGGCGGAGATCNAAGCCTTCATCAAGCTGTAC	363
OY	432	-ACTGGGATCTGCGCTCAGGCGGCTTACGATGCTATGAGG-CTGGCTCAACGTGGAAAGA	489
Db	364	AACCTGGGACTCGCCGACGAGANTGGACCGGCGGTACCGCGGCGCTCTGAACCGCACANA	423
OY	490	GGTCAGCTGACCTTGAACGGTATGCGAG	519
Db	424	GTTCCCGCTGCACTTGGCCCGGTACGCGCG	453

RESULT	11
AY106991	
LOCUS	
DEFINITION	AY106991 Zea mays PC0088410 mRNA sequence.
ACCESSION	
VERSION	AY106991.1 GI:21210069
KEYWORDS	
SOURCE	HTC.
ORGANISM	Zea mays
	Zea mays

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 1182)	Hainey, C. F., Dolan, M., Miao, G. H., Vogel, J. M., Whitsitt, M. S., Arthur, L. W., Hanafey, M., Morgante, M. and Tingey, S. V.	Matze Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes

UNPUBLISHED (2002)
2 (bases 1 to 1182)
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Coe, E. H.
Direct Submission
Submitted (25-Apr-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDBs and may be found by BLAST
searching at MSU, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDBs:
www.zmdb.iastate.edu.

```

FEATURES
source
location/Qualifiers
1..1182
/organism="Zea mays"
/mol_type="rRNA"
/db_xref="MaizeDB:635002"
/db_xref="taxon:4577"
/clone_11b="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
configs to seed DuPont configs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of Bacs in conjunction with the Maize
Mapping Project"

```

	ORIGIN	
	Query Match	8.8% Score 175; DB 11; Length 1182;
	Best Local Similarity	54.8%; Pred. No. 5,9e-21;
	Matches 440; Conservative	0; Mismatches 345; Indels 18; Gaps 4;
QY	77 GCTCTGCCCAACGACTTGAATGGAGCCTTCGCACGCCCGCCCTACAGATCGAAGCGCC	136

Db	205	GCCTTCCGGACGCGCTTCGTTTCGGGACGCGCGCTCGCGCTACAGGTCGAGGGGATG	264
Oy	137	GTCAAGAAGTGGCCCGCGCCCGCTCATTTGGGACACGTACTGCCACTCGAGCCATCG	196
Db	265	GCCAGACGCGCGCGCGCGCGCCCGACGATCTGGGACCGCTTCATAGAGGTTCCCGGACC	324
Oy	197	CGCACCAAGCGCGGCAACGGGCGATGTTGGGATCGTACACACGCGTACGATGAGGAC	256
Db	325	ATCCCTTACAAATGCTACCGCTGACGTACCGGTACGAGTATCATCGGTACAGGAAGAT	384
Oy	257	TTTGATCTCTTGACCAAGTACGCGGCAAAAGGCTACCGCTTCTCTTGCTGGTCCGG	316
Db	385	GTGAACATATGAAAGAACATGGCGTTTGATGCGTACCGGTTTTCGATCTCTGTGTCAGG	444
Oy	317	ATCATTCCTCCCTCGGCGGACGAGCTGATCCGCTCAACAGGAGGGAATTGATTTACAGC	376
Db	445	ATTTTCCAGATGGAACTGGCAAG-----GTAAACAGGAAGAGTGGATTTACTACAC	498
Oy	377	AAACTGATTCAGCGCTTGTGAGGCGGGGATATCAGCGCTTGGGTGACTTTGTACACTGG	436
Db	499	AGGCTCATATGATTCATGCTCCACAGCAAGTATCGCGGTATGCAAACTCTTACATTAT	558
Oy	437	GATCTGCTCAGGCGGCTTCACGATCGCTATGAGAGGTGCTCAACGTGGAAGAGTCCAG	496
Db	559	GACCTCCCATTTGGACATCCACATGAAACAGTACCTGGGCTGGCTTACCC---CAAAATTGTG	615
Oy	497	CTGACATTTCGAGCGGTATGCGAGGTTGTGCTTTGTAAGCTTTTGGGGAACGAGTCAGAAC	556
Db	616	GAGCGCTTTGGAGACTACGCGCGAGTCTTGCTTCCACGCGTTCCGAGACAGGGTGAAGAC	675
Oy	557	TGATATCAACATCAACGAMCCCTGGATTCAAGGCATCTATGATATGCCACCGCAGCAAC	616
Db	676	TGCTTTTACTTTCACGAGCGGAGGTGGCTGGCTCTTGAGCTACGACATATGCTTGCAC	735
Oy	617	GCCCCGGGACGAGCAGCATTTAACAMGACTCCACGAGGGCAACACTGCCACTGAGCCG	676
Db	736	GCACCGGGAAGTGT-----TTCCGGGTGCCCCGCGGAGGCAACTCACACGAGACCG	789
Oy	677	TGGCTCGCTGAAAGGCCCAAGATCATGAGCCATGCCCGCGCGCTGCGCTTCACGACG	736
Db	790	TACCTTTCGACACCATCTCATCTCTTTCATCATGCACTCGGTCAGAGCATACCGGAC	849
Oy	737	GACTTTGGCCCTTGCAAAAGGGCCAAATGGGACTCTGCTCAACGGCGACTACTATGAG	796
Db	850	AAGTATCAGCTTCCACGAAGGGGAGAAATTGGAATTTCTCTGGATTTCGTGTGTACGAA	909
Oy	797	CCCTGGGACAGCAATGAGCTCGGGAACAAGAGGCTCTAGCGAGGATGGAATTTAC	856
Db	910	CTTT---TCAGCGACAGCAATGCGGACACGAGCTGACACAGCGAGCCAGGACTTCCAC	966
Oy	857	ATTGCTGCTGTTTGGCAATCCCAT	879
Db	967	CTAGGCTGGTTCTTGAACCCCAT	989

RESULT 12					
LOCUS	CF880605				
DEFINITION	CF880605	662 bp	mRNA	linear	EST 31-OCT-2003
ACCESSION	trich082xh20_b1	Tr.reesl mycelial culture,			Version 6 October 2003
VERSION	Hypocrea jecorina cDNA clone	trich082xh20,			mRNA sequence.
KEYWORDS	CF880605.1	GI:38135287			
SOURCE	Est.				
ORGANISM	Hypocrea jecorina (anamorph: Trichoderma reesei)				
	Hypocrea jecorina				
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
	Hypocreomycetidae; Hypocreales; Hypocricaceae; Hypocrea.				
REFERENCE	1 (bases 1 to 662)				
AUTHORS	Dienes, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,				
	Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and				
	Dean, R.A.				
TITLE	Analysis of the protein processing and secretion pathways in a				
	Trichoderma reesei EST dataset				

JOURNAL Unpublished (2003)

COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph_dean@ncsu.edu

FEATURES

Seq primer: Lt-F1 primer.
Location/Qualifiers

1. 662
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric082xh20"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October 2003"
/note="Vector: PREP3J; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 8.6%; Score 171.8; DB 14; Length 662;
Best Local Similarity 59.4%; Pred. No. 2.6e-20;
Matches 366; Conservative 0; Mismatches 238; Indels 12; Gaps 4;

```

QY 646 CTCACCGAGGCGAACACTGCGACCTGCGCTGCTGCGAAAGGCCGATCATGAG 705
Db 52 CGCCCCCGCGCGAGACACCTCGAGCCGCGAGCCTGCGCCCAACATCTGTCGC 111
QY 706 CCATGCCCCCGCGCGCTGCGCTTACAGCAGGAGACTT--TCGCCCTCGCAAAAGGGCCA 762
Db 112 CCACGCGCGCGCGCTCAAGGGGTACCGGAGAGACTTCAAGCCCGCAGCGCGCA 171
QY 763 GATCGGATCTTCCTCAACCGGCACTTATGAGCCTCGGACAGCAATGAGCCTCGGGA 822
Db 172 GATCGGATCTTCCTCAACCGGCACTTATGAGCCTCGGACAGCAGCGCGCGCA 231
QY 823 CAAGAGGCTGCTGAGGAGCGAGATTTCACTTGGCTGGTTTGCATCTCCATCTT 882
Db 232 CAAGAGGCTGCTGAGGAGCGAGATTTCACTTGGCTGGTTTGCATCTCCATCTT 291
QY 883 CTGAAGAGGACTATCCAGAGCAGTGAAGAGCAGCTGGCGAGAGGCTTCCAGCCT 942
Db 292 CTT---GGGCGACTACCGCGGTGATGCGAAGAGCTGGGCGACCGGCTCGGACTT 348
QY 943 CACTCCCGCGACTTTGGCATCTTCATGTCGCGAGAGACCGACTTTACGGCATGAATTA 1002
Db 349 TACGCGCGAGAGCGCGCTCTGCTCA---CGGCTCCAAAGCATTTTACGGCATGAACA 405
QY 1003 CTACACATCCGAGTTTCGCGCGCACTAGACGATCCCGCTCCGAGACGACATCTCTCG 1062
Db 406 CTACACGTTCAATCAATCTCCGACCGAGAGTCCGCGCTCCGCGACGACCGCTG 465
QY 1063 CGCATTCATAGACACAGAGAAATAAGACGAGCGCCGCTTGGCGAGAGACGCGCT 1122
Db 466 CAAGGTGACGCTGCTTTCACCAACAGCAGGCAATGATCGGCCCGCAGACGAGTC 525
QY 1123 CGCGTGGCTGCGCTTCGCGCGCAATGTTCCGGAAGCATTCGCGCGGCTTACGCGCT 1182
Db 526 CCGCTGGCTGCGCGCTTCGCGCGCAATGTTCCGGAAGCATTCGCGCGGCTTACGCGCT 585
QY 1183 GTACGCGAA--GCCATCTCATACCGAGAACGAGATGCCGCTCGCGTGGAGAGGAA 1239
Db 586 GTACGCGTACCGCGCGCATCTAGAGAGAGAACGACACAGACTCAAGGGCGAGAGCGA 645
QY 1240 CATGACGTGAGAGAG 1255
Db 646 CTTGCCCAAGAGAGAG 661

```

RESULT 13

CB907493 752 bp mRNA linear EST 02-JUL-2003

LOCUS 752 bp mRNA linear EST 02-JUL-2003

DEFINITION trico82xh20 T.reesei mycelial culture, Version 3 april Hypocrea

ACCESION jecorina cDNA clone trico82xh20, mRNA sequence.

VERSION CB907493

KEYWORDS CB907493.1 GI:30122151

SOURCE EST.

ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)

REFERENCE Hypocrea jecorina

AUTHORS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

1 (bases 1 to 752)

Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,

Dunn-Coleman, N.S., Goedegebuur, P., Houfek, T.D., England, G.J.,

Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,

Olivares, H.A., Teunissen, P.J., Yao, J., and Ward, M.

Transcriptional regulation of biomass-degrading enzymes in the

filamentous fungus Trichoderma reesei

J. Biol. Chem. 278 (34), 31988-31997 (2003)

JOURNAL MEDLINE 22603314

PUBMED 12788920

COMMENT Contact: Pamela K. Foreman

Genencor Int'l.

925 Page Mill Road, Palo Alto, CA 94304, USA

Tel: (650) 846-7635

Fax: (650) 621-7817

Email: Pforeman@genencor.com

Seq primer: Lt-F1 primer.

Location/Qualifiers

1. 752

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/strain="QM6a"

/db_xref="taxon:51453"

/clone="trico82xh20"

/dev_stage="mycelia"

/note="Vector: PREP3J; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 8.6%; Score 171.8; DB 14; Length 752;
Best Local Similarity 59.4%; Pred. No. 2.5e-20;
Matches 366; Conservative 0; Mismatches 238; Indels 12; Gaps 4;

```

QY 646 CTCACCGAGGCGAACACTGCGACCTGCGCTGCTGCGAAAGGCCGATCATGAG 705
Db 142 CGCCCCCGCGCGAGACACCTCGAGCCGCGAGCCTGCGCCCAACATCTGTCGC 201
QY 706 CCATGCCCCCGCGCGCTGCGCTTACAGCAGGAGACTT--TCGCCCTCGCAAAAGGGCCA 762
Db 202 CCAGGCGCGCGCGCTCAAGGGGTACCGGAGAGACTTCAAGCCCGCAGCGAGCGCA 261
QY 763 GATCGGATCTTCCTCAACCGGCACTTATGAGCCTCGGACAGCAATGAGCCTCGGGA 822
Db 262 GATCGGATCTTCCTCAACCGGCACTTATGAGCCTCGGACAGCGCGCGCGCGCA 321
QY 823 CAAGAGGCTGCTGAGGAGCGAGATTTCACTTGGCTGGTTTGCATCTCCATCTT 882
Db 322 CAAGAGGCTGCTGAGGAGCGAGATTTCACTTGGCTGGTTTGCATCTCCATCTT 381
QY 883 CTTGAAGAGGACTATCCAGAGGAGTGAAGAGAGTGGGCGAGAGGCTTCCAGCCT 942
Db 382 CTT---GGGCGACTACCGCGGTGATGCGAAGAGTGGGCGAGCGCTTCCACCTT 438
QY 943 CACTCCCGCGACTTTGGCATCTTCATGTCGCGAGAGACCGACTTACGGCATGAATTA 1002
Db 439 TACGCGCGAGAGCGCGCTCTGCTCA---CGGCTCCAAAGCATTTTACGGCATGAACA 495
QY 1003 CTACACATCCGAGTTTCGCGCGCACTAGACGATCCCGTCCCGAGACGAGACTATCTCG 1062

```

Db 496 CTACAGCTCCACATACATCCGACCGAGCTCCGCTCCGCGACGACACCGTGG 555
Qy 1063 CGCATTCATGACACACAGAGATAAGACGGAGCCGCTGGGAGAGAGCGGCTT 1122
Db 556 CAAGCTGACGACTCTTACCAACAGAGGCACTGATCGGCGCCGACGACGATC 615
Qy 1123 CGCTGAGCTGCTCTGCGCCGACATGTTCCGAAAGCATCTGCCGAGGTATACGCTT 1182
Db 616 CCCCTGGCTGCGCCCTGTGGCCGCGCTTCCGCACTTCTGTGTGATCAGNCAGAG 675
Qy 1183 GTACGGCAA--GCCATCTACATCAACGAGATGCGCTGCTGAGAGAGAA 1239
Db 676 GTACGGCTACCGCCCATCTACGTGACGAGAACGACGACATCAAGGCGAGAGCA 735
Qy 1240 CATGACGTGCGAGAG 1255
Db 736 CTTGCCCAAGAGAG 751

RESULT 14
BE777259 650 bp mRNA linear EST 20-SEP-2000
LOCUS MY-26-C-10 PinfestansMY Phytophthora infestans cDNA, mRNA sequence.
DEFINITION BE777259
ACCESSION BE777259 GI:10230914
VERSION BE777259.1
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
1 (bases 1 to 650)
Kamoun S., Harber P.T., Sobral B.W.S., Nuss D. and Govers F.
Initial assessment of gene diversity for the oomycete pathogen
Phytophthora infestans based on expressed sequences
Fungal Genet. Biol. 28 (2), 94-106 (1999)
20056376
10587472
Contact: Govers F
Laboratory of Phytopathology
Wageningen University
Bilthoven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands
Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyco.wau.nl.

FEATURES
source
1. 650
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="DDR7602, A1 mating type"
/db_xref="taxon:4787"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DH5-alpha"
/clone_lib="PinfestansMY"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Total
RNA was isolated from mycelium of P. infestans DDR7602
cultured for 4 weeks in synthetic medium. EST clones were
named by their position in the microtiter plate, preceded
by the prefix MY (for mycelial) and the successive number
of the microtiter plate (e.g. MY-06-A-04)."

ORIGIN
Query Match 8.5%; Score 169.8; DB 10; Length 650;
Best Local Similarity 58.1%; Pred. No. 5.8e-20;
Matches 319; Conservative 0; Mismatches 224; Indels 6; Gaps 1;

Qy 81 TGCCCAACGACTTGATGGAGGCTTGCACAGCGCGCTACGATCGAAGGCGCGCTCA 140
Db 58 TCCCGAAGACTTATATGTGGGACATGCCACTGCTGTGACCAAGTGGAGGTGACGA 117
Qy 141 AAGAGGTGGCGCGCGCGCTTCCATCTGGGACAGTACTGCCAATTGAGACCATCGGCA 200
Db 118 ACGAAGTGGACCGCGGACAGCATCTGGACGCTTCTCTGGACACCGGCAAGATCG 177

Qy 201 CCAACGGCGCAACGGCGATGTGGCTTGGCATCTACCAACCGCTACGATGAGACTTTG 260
Db 178 TGAATGAGACACGGCGGAGAGAGCGGTGATATATATCGCTACAGAGAGACTGC 237
Qy 261 ATCTCTTGAACCAATGACGGCGCAAGGCTTACCGCTTCTCTTGTGTGTGCGGATCA 320
Db 238 AACTTATGAAGAAGATGGAGACTCAAGCTTACCGTCTCTTCATCGCTTGGCCGCGATCA 297
Qy 321 TTCCCTCGCGGACAGCTGATCCGCTCAACGAGAGAGGAATTGATTACAGCAAC 380
Db 298 TCCCGCAGAG-----TCGGCGAGTGAACAGAGAGCGCTGAGTTCTACAAACACC 351
Qy 381 TGATTACGCCCTGTGAGAGCGGGGTATACAGCGCTTGGGTAATTGTATCACTGGGATC 440
Db 352 TCACTAATAGACTACTGGCCACAGACATCACGCCCTGTTACGCTTACATTTGGATT 411
Qy 441 TGCTGAGCGCTTTCAGATCGCTATGAGAGCTGAGCTCAACGTGAAGAGTCCAGCTGG 500
Db 412 TACCTTAGCTGTGCAAAACGAGTATGACGGTTGGCTTGGTGGCAGATCATCAAGACG 471
Qy 501 ACTTTAGCGGTATGCGAGGTGTGCTTTGAAAGCTTTGGGACCGAGTCCAGACTGGA 560
Db 472 CCTTCGTGCAATPACGCTCGCTGTGCTTCCAGCGCTTGGAGACCGTGTGACCACTGGT 531
Qy 561 TCACCATCAACGACCCCTGATTCAGGCCATCTATGAGATATGCGACCGGACGACGCC 620
Db 532 TGACGTGAATGAGCGGTGTGCTCCGCTTCTTGAGGCTATGAGATGATGTTCAATGCTN 591
Qy 621 CGGGCAGGA 629
Db 592 CAGAGCA 600

RESULT 15
AJ436490 613 bp mRNA linear EST 15-MAR-2002
LOCUS AJ436490
DEFINITION AJ436490 S00007 Hordeum vulgare cDNA clone S0000700006A08F2, mRNA
sequence.
ACCESSION AJ436490
VERSION AJ436490.1 GI:19524942
KEYWORDS EST.
SOURCE Hordeum vulgare
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 613)
Saren A.-M., Tanskanen J., Paulin L. and Schulman A.H.
Barley EST's
Unpublished (2002)
Contact: Schulman AH
Institute of Biotechnology
University of Helsinki
P.O.Box 56 (Valkinkari 6A), University of Helsinki FIN-00014,
Finland.

FEATURES
source
1. 613
/organism="Hordeum vulgare"
/mol_type="mRNA"
/db_xref="taxon:4513"
/clone="S0000700006A08F2"
/dev_stage="Shoot"
/clone_lib="S00007"
/note="2,-3,-4-days after germination"

ORIGIN
Query Match 8.2%; Score 162.6; DB 9; Length 613;
Best Local Similarity 57.2%; Pred. No. 1.1e-18;
Matches 338; Conservative 0; Mismatches 244; Indels 9; Gaps 2;

Qy 204 ACGGCCCAACGGCGATGTGGCTTGGATCACTACCAACGCTTACGATGAGACTTGTGATC 263

```
Db      19  ACGGCAACCACTGGATACAGCAAAACGACTTCTATCATCGCTACAGAGAGATCTGAAGC 78
Qy      264 TCTTGACCAAGTACGGCGCAAAAGGCTTACCGCTTCTCCTTGGTGGTGGATCATTC 323
Db      79  TGGTACCTGACATGAATGAATGACCGCTTCCGCTTCTCCCTTGGCATGAGACAGGATCCTTC 138
Qy      324 CCTCGCGCGGAGGCTGGATCCCGTCAACGAGAGGAAATTGAGTTTACAGCAAACTGA 383
Db      139 CAATGGACCATATGCGCGAGGAAATCAAAAGCAGAGATGATTTCTACAAACAGCCTCA 198
Qy      384 TTGACGCCCTGTGTGAGCGGGGATTCACGCTTGGGTGACTTTGTACCACTGGGATCTGC 443
Db      199 TCGACGAGGTTTGTGGCTAGAGGGCTGATGCTTTCATCAATGTTCCACTTCGACACCC 258
Qy      444 CTCAGGCGCTTTCAGGATCGCTATGAGAGCTGGCTCAAGTGAAGAGATCCAGCTGACT 503
Db      259 CCGAGGCTCTGAGGACAAATACGGAAGCTTCTTGAGC--GACACATCGTAAAGGACT 315
Qy      504 TTGAGCGGATATGCGAGGTTGTCTTGAACGTTTGGGGAACGATCCAGAACTGATCA 563
Db      316 ACGTGAGATACGCGGAGCTGTGCTTCAAGCTGTTGCGCGACCGGGTGAAATTCGACCA 375
Qy      564 CCATCAACGAMCCTTGATTCAGGCCCATCTATGATATGCCACTGGCAGCAACGCCCGG 623
Db      376 CTTTCAACGAGCCCATGATGTTCTGCGCTTCGCGCTACGGCACCGGACCTTGGCCCCGG 435
Qy      624 GCAGGAGC-----AGCATTAACAAAGCACTCCACGAGGGCAACACTGCCACTGAGCGT 677
Db      436 GCGGCTGTGCGCCGTACGTCTCCAAAGGCTGCGGCGCCGCGGAGCTCTCCACGAGGCTCT 495
Qy      678 GGCCTGCTGAAAGGCCCAATCATAGACCAATGCCCGCGCGGTGGCCGTCTACAGCAGGG 737
Db      496 ACATCGCGGCGACACACTCTCTCATGCGCCACGCGGAGCGGTGCACCTGTACCGCACCA 555
Qy      738 ACTTTCGCCCTTCGCAAAAGGCCCAATCGGCATCTCGCTCAACGGGACT 788
Db      556 GGTACCAAGCGGCGCAGCGGGGCCAGATCGGCATCGTCAATATCCCACT 606
```

Search completed: March 28, 2004, 23:31:35
Job time : 5351.5 secs

Oliveres, H. A., Teunissen, P. J. M., Yao, J. and Ward, M.
Transcriptional Regulation of Biomass-Degrading Enzymes in the
Filamentous Fungus *Trichoderma reesei*
J. Biol. Chem. (2003) In press
2 (bases 1 to 1455)
Foreman, P. K., Brown, D., Dankmeyer, L., Dean, R., Diemer, S.,
Dunn-Coleman, N. S., Goedegebuer, F., Houfek, T. D., Mitchell, T.,
Oliveres, H. A., Teunissen, P. J. M., Yao, J. and Ward, M.,
Direct Submission
Submitted (23-APR-2003) Genencor Intl., 925 Page Mill Road, Palo
Alto, CA 94304, USA

FEATURES
source
1..1455
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="Om6a"
/db_xref="taxon:51453"
/note="anamorph: *Trichoderma reesei*"
1..1455
/gene="cel1b"
1..1455
/gene="cel1b"
/note="putative beta-glucosidase"
/codon_start=1
/product="Cel1b"
/protein_id="AAP57758.1"
/db_xref="GI:31747170"
/translation="MPESLALPNDPEMFATAYOISGAVKEGRGSPIMDTCHLP
SRINGANGDVACDHYRDEPDLITKGAAYRPSLSMSLIIPLGRLDVEEGEIE
FYSKLIDALLRGITPPWTLTHMDLPOLHRYGMLNVEVDLFEYALDCEKRG
DRVQNWITNKPWIOALTYGATGNSAPERSINSHTEGNTATPEMLAKQINSHAR
AVAYSRDRFSPQKQIGISLNGDIYEPEDNEPRDKEAERMEFHIQWNPFIK
KDYEMKQKQGERLPAITPADFAILNAGETDFGMVYTSQFARHLDGPEPDYIG
AIHHEQNKDGSPIVSEESGLAMLRSCPMFKHILARYGLKGFYITENCCPRGEE
NMTCBEAVNDPFRIRYFSDHLSISKAITDGVVVKGFAMALLDNEMSDGYPRRG
VTPTDYTLTKRTPKKSALVLDMPAARQVRVVA"

ORIGIN

Query Match 73.0% Score 1454.2; DB 8; Length 1455;
Best Local Similarity 100.0%; Pred. No. 8.8e-285;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 ATGCCGAGTGGCTNAGCTCTGCGCCCAACGACTTTGATGGGCTTGCAACGCGCGCTTAC 121
DB 1 ATGCCGAGTGGCTNAGCTCTGCGCCCAACGACTTTGATGGGCTTGCAACGCGCGCTTAC 60

QY 122 CAGATCGAAGGCGCGCTCAAGAAAGGTGGCGCGCGCGCTTCATCTGGGACACGATCTGC 181
DB 61 CAGATCGAAGGCGCGCTCAAGAAAGGTGGCGCGCGCGCTTCATCTGGGACACGATCTGC 120

QY 182 CACCTGAGCAATGCGCGCACCAACGCGCGCAACGCGCGATGAGCTTGCGATCACTACAC 241
DB 121 CACCTGAGCAATGCGCGCACCAACGCGCGCAACGCGCGATGAGCTTGCGATCACTACAC 180

QY 242 CGCTACGATGAGCACTTGATCTCTTGACCAAGTACGCGCGCAAGGCGCTTACCGCTTCTCC 301
DB 181 CGCTACGATGAGCACTTGATCTCTTGACCAAGTACGCGCGCAAGGCGCTTACCGCTTCTCC 240

QY 302 TTGTCTGTGTGCGCGGATCAATCCCTCGCGCGGAGCTGATCCCGTCAACGAGAGGAGA 361
DB 241 TTGTCTGTGTGCGCGGATCAATCCCTCGCGCGGAGCTGATCCCGTCAACGAGAGGAGA 300

QY 362 ATTGATGTTTACGACAACTGATGACGCGCTGTTGAGCGCGGATACAGCGCTTGAGTG 421
DB 301 ATTGATGTTTACGACAACTGATGACGCGCTGTTGAGCGCGGATACAGCGCTTGAGTG 360

QY 422 ACTTTGATCAACCTGGGATCTGGCTCAAGCGCTTACAGATGCTATGAGAGCTGGCTAAC 481
DB 361 ACTTTGATCAACCTGGGATCTGGCTCAAGCGCTTACAGATGCTATGAGAGCTGGCTAAC 420

QY 482 GTGAGAGAGGTCAAGTGAATTTGAGCGGATACGAGAGTGTGCTTGAACGTTTTGGG 541
DB 421 GTGAGAGAGGTCAAGTGAATTTGAGCGGATACGAGAGTGTGCTTGAACGTTTTGGG 480

QY 542 GACCGATCCAGAACTGGATACCATCAACAGACCTTGATTTAGGCGCATCTATGATAT 601
DB 481 GACCGATCCAGAACTGGATACCATCAACAGACCTTGATTTAGGCGCATCTATGATAT 540

QY 602 GCCACCGGACAGCAAGCGCGCGGACAGAGCATTTAACAAGCACTCCACCGAGGAGCAAC 661
DB 541 GCCACCGGACAGCAAGCGCGCGGACAGAGCATTTAACAAGCACTCCACCGAGGAGCAAC 600

QY 662 ACTGCCACTGAGCGCTGCTGCTGGAAGAGCCCAAGTCAATGAGCAATGCGCGCGGTG 721
DB 601 ACTGCCACTGAGCGCTGCTGCTGGAAGAGCCCAAGTCAATGAGCAATGCGCGCGGTG 660

QY 722 GCGCTCAACAGCAGGAGCTTTCGCGCGCAAAAGGCGCAAGTCCGATCTGCTCAAC 781
DB 661 GCGCTCAACAGCAGGAGCTTTCGCGCGCAAAAGGCGCAAGTCCGATCTGCTCAAC 720

QY 782 GCGCATCTATGAGCGCTGCGGACAGCAATGAGCTTCGGAACAAGAGGCTGTGAGCGA 841
DB 721 GCGCATCTATGAGCGCTGCGGACAGCAATGAGCTTCGGAACAAGAGGCTGTGAGCGA 780

QY 842 CGGATGGAATTTCACTTGGCTGCTTGGCAATCCCATCTTCTTGAAGAAAGCATATCCA 901
DB 781 CGGATGGAATTTCACTTGGCTGCTTGGCAATCCCATCTTCTTGAAGAAAGCATATCCA 840

QY 902 GAGAGCATGAAGAAGCAGCTGGGCGAGAGGCTTCCAGCGCTCCAGCGCGGACTTTGCC 961
DB 841 GAGAGCATGAAGAAGCAGCTGGGCGAGAGGCTTCCAGCGCTCCAGCGCGGACTTTGCC 900

QY 962 ATCTCTAATGCGGAGAGACCGAATTTCTACGCGCATGAATTTACTACATCTCCAGTTGCGG 1021
DB 901 ATCTCTAATGCGGAGAGACCGAATTTCTACGCGCATGAATTTACTACATCTCCAGTTGCGG 960

QY 1022 CGCCACTTAAAGCTGCTGCTGCTGCGGAGAGGACATATCTGCGCGCATTCATGAGCAACG 1081
DB 961 CGCCACTTAAAGCTGCTGCTGCTGCGGAGAGGACATATCTGCGCGCATTCATGAGCAACG 1020

QY 1082 GAGAAATGAAGCAGGACCGCGCTTGGCGAGAGAGGCGCTCGCTGCGCTGCGCTCTGC 1141
DB 1021 GAGAAATGAAGCAGGACCGCGCTTGGCGAGAGAGGCGCTCGCTGCGCTGCGCTCTGC 1080

QY 1142 CCGGACATGTTCCGGAAGCATCTTGGCGCGGATGTAAGCGCTGTACGGAAGCCCATCTAC 1201
DB 1081 CCGGACATGTTCCGGAAGCATCTTGGCGCGGATGTAAGCGCTGTACGGAAGCCCATCTAC 1140

QY 1202 ATCACCGAAGAGGAGTCCCGCTGCGCTGAGAGGAGAAATATGACCTGCGAGAGGCGCTG 1261
DB 1141 ATCACCGAAGAGGAGTCCCGCTGCGCTGAGAGGAGAAATATGACCTGCGAGAGGCGCTG 1200

QY 1262 AACGACCCCTTCCGATCCGATCTTGTGACTGCACTTGATTTCAAGGCGCATTT 1321
DB 1201 AACGACCCCTTCCGATCCGATCTTGTGACTGCACTTGATTTCAAGGCGCATTT 1260

QY 1322 ACCCAGAGGCGGCTGCTGCTCAAGGAGTACTTGTGCTGAGGCTGTGATTAATTGGA 1381
DB 1261 ACCCAGAGGCGGCTGCTGCTCAAGGAGTACTTGTGCTGAGGCTGTGATTAATTGGA 1320

QY 1382 TGGTCAATGAGGCTAGCGGACCGCATTTGCGGCTGCAAGTCAACGATCCCAAG 1441
DB 1321 TGGTCAATGAGGCTAGCGGACCGCATTTGCGGCTGCAAGTCAACGATCCCAAG 1380

QY 1442 CGCAGCCCAAGAAAGTCTGCGCTGCTCAAGAGCATTTTGGCGCGCGGAGAGGTT 1501
DB 1381 CGCAGCCCAAGAAAGTCTGCGCTGCTCAAGAGCATTTTGGCGCGCGGAGAGGTT 1440

QY 1502 AAAATGGCGCATTA 1516
DB 1441 AAAATGGCGCATTA 1455

RESULT 2
AB003109
LOCUS AB003109 1900 bp DNA linear PLN 25-DEC-2002

DEFINITION Humicola grisea var. thermoides bg14 gene for beta-glucosidase, complete cds.

ACCESSION AB003109

VERSION AB003109.1 GI:4249559

KEYWORDS beta-glucosidase.

SOURCE Humicola grisea var. thermoides

ORGANISM Humicola grisea var. thermoides
Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.

REFERENCE
1 Takahima, S., Nakamura, A., Hidaka, M., Masaki, H. and Uozumi, T. Molecular cloning and expression of the novel fungal beta-glucosidase genes from Humicola grisea and Trichoderma reesei J. Biochem. (1999) in press

JOURNAL Uozumi, T.
2 (bases 1 to 1900)

REFERENCE Direct Submission
Submitted (09-MAR-1997) Takeshi Uozumi, The University of Tokyo, Department of Biotechnology, Faculty of Agriculture, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan (E-mail: uozumi@mb.d.t.u-tokyo.ac.jp, Tel: 03-5684-0387, Fax: 03-5684-0387)

TITLE Location/Qualifiers

JOURNAL 1. 1900

FEATURES /organism="Humicola grisea var. thermoides"
source /mol_type="genomic DNA"
/sub_species="IFO9854"
/db_xref="taxon:5528"
join(289..332,437..1823)
/gene="bg14"
join(289..332,437..1823)
/gene="bg14"
/codon_start=1
/product="beta-glucosidase"
/protein_id="BA074958.1"
/db_xref="GI:4249560"
/translation="MSLPDPKFWGFATPAVQIEGVSNEDEGSPSINDTFCAIEGKIAD
GSSGAAVACSRYKTKEDIALKELGANSYFSISWSRIPLGRNBPINOKGIDHYK
FVDLLIAGITPEITLFLHMDLPDLKRYGFLNKEFPADFPNVAIPNKAIPCKH
WTFENPMSALIGVNTGYPAPGTSRSPGSDSREPMWIGHNILLIHAARAVAY
REDKFPDGGEGITLNGDALTPMDPEPDADICADKRIEPAISWADPDIYRKYSDS
MRQOLDRLPEFTPEVALYKGSNDPYGNMHTYANTYIKHTGCPPEDDFLGNLETFY
NKYDGCIGPETOSFWLPHAGFRDLNLSKRYGPKYVTGNGSLKENDMPLEO
VLEDPRVKYFNDYVRAAATAVAEDGNCNMGYLAWSLIDNFEWAEVETRFVTYVDY
ANDQKRYPKSAKSLKPLPDLIRKE"

ORIGIN

Query Match 24.2%; Score 481.8; DB 8; Length 1900;
Best Local Similarity 61.7%; Pred. No. 2e-87;
Matches 854; Conservative 1; Mismatches 513; Indels 15; Gaps 5;

108 CAACGGCCGCTACAGATGGAAGGCGCGCTCAAGAGAGGTGGCCGCGCCGCTCATCT 167
Db 427 CAATCAAGCTACAGATGGAAGGCTCCGCTCAAGAGAGGTGGCCGCGCCGCTCATCT 486
Qy 168 GGGACAAGTACTGCACTGGAGCCATCGGCAACAAGCGCGCAACGGAGTGGCTT 227
Db 487 GGGACAAGTACTGCACTGGAGCCATCGGCAACAAGCGCGCAACGGAGTGGCTT 546
Qy 228 GCGATCACTACCAAGCTACGATGAGACTTTGATCTTTGACCAAGTACGGCGCAAG 287
Db 547 GCGATCACTACCAAGCTACGATGAGACTTTGATCTTTGACCAAGTACGGCGCAAG 606
Qy 288 CCTACCGCTTCTCTGTGCTGGTGGCGGATCAATCCCTCGGCGGCGAGCTGATCCCG 347
Db 607 CGTACCGCTTCTCTGTGCTGGTGGCGGATCAATCCCTCGGCGGCGAGCTGATCCCG 666
Qy 348 TCAACGAGAGGAGATTGAGTTTACAGCAACTGATTGACGCGCTTTGAGGCGGGGTA 407
Db 667 TCAACGAGAGGAGATTGAGTTTACAGCAACTGATTGACGCGCTTTGAGGCGGGGTA 726
Qy 408 TCAACGAGGAGGAGATTGAGTTTACAGCAACTGATTGACGCGCTTTGAGGCGGGGTA 467
Db 727 TCAACGAGGAGGAGATTGAGTTTACAGCAACTGATTGACGCGCTTTGAGGCGGGGTA 786

Qy 468 GAGCTGAGCTCAACGTGGAAGAGTCCAGCTGACTTTGAGCGGTATGCGAGGTGTGCT 527
Db 787 GCGGCTTCTCCACAGAGAGAGAGTTCGCGCGCACTTTGAGAACTAGCCCGCATCATGT 846
Qy 528 TTGAAGCTTTGGGGAACCGAGTCCAGACTGTATCAACATCAACGAGCCCTGATTCAG 587
Db 847 TCAAGGCAATCCCAAGT---CAAGCACTGATCACTTCAACGAGCCCTGTGTCTCCG 903
Qy 588 CCATCTATGATATGCAACCGGAGAGAGCCCGCGGAGAGACACATTAACAGACT 647
Db 904 CCATCTCTGCTCAACAGAGGCTACTTGGCCCCGACACACTTGGAGCCGAGCAAGT 963
Qy 648 CCAACGAGGCAACACTGCACTGAGCCGTGCTGCTGGAAGAGCCAGATCATGAGCC 707
Db 964 CCCCCCTCGGCAACAGCCCGGAGCCCTGATTTGTCCGCCCAACATCTCTATGCCC 1023
Qy 708 ATCCCGCGCGCTGCGCTTCTACAGAGGACTTTGCGCCCTCGCAAGAGGCGCAATCG 767
Db 1024 ACCCGCGCGCTCAAGGCTTACCGGAGGACTTCAAGCCACCCAGGCGCGAGATCG 1083
Qy 768 GCATCTCGCTCAACGGCGACTATGAGCCCTGGGAGAGCAATGAGCTCGGAGCAAG 827
Db 1084 GCATCAACCTGAAGCGCGAGCCACCTCTCGGAGCCCGAGACCCGCGAGACTTG 1143
Qy 828 AGGCTGCTGAGCAGCGAGTGAATTTCACTTGGCTGTTGGCAATCCCATCTTCTTGA 887
Db 1144 AGGCTGCTGAGCAGCGAGTGAATTTCACTTGGCTGTTGGCAATCCCATCTTCTTGC 1203
Qy 888 AGAAGACTATCCAGAGAGATGAAAGAGAGCTGAGGCGAGAGGCTTCCAGCCCTCACTC 947
Db 1204 GCAAGTAC---CCGACTGATGCGAGAGAGCTGAGGCGAGCCGCTGCGGAGTCACTC 1260
Qy 948 CCGCGAGCTTGGCATCTTCAATGCGCGAGAGAGCGACTTTCAGGAGATTAATTAACA 1007
Db 1261 CCGAAGAGTGGCGCTGCTCA---GGGCTCCAGCACTTTCAGGAGATTAATTAACA 1317
Qy 1008 CATCCAGTTCGGCGCCCACTTGAAGTCCGCTCCCGAGAGAGACTTCTCGCGCA 1067
Db 1318 CGGCAACATCAAGCAAGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1377
Qy 1068 TCCATGAGCAACGAGAAATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1127
Db 1378 TCGAGAGCTCTTTTACCAAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1437
Qy 1128 GGTGCGCTCTGCGCGGAGATGTTCCGAGAGATTCGCGCGGAGTACGCGCTGATCG 1187
Db 1438 GGTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1497
Qy 1188 GCAAGGCC---ATCTATCATCAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1244
Db 1498 GCTACCCCAAGATCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1557
Qy 1245 CCGTGGAGAGGCGGCTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1304
Db 1558 CCTCGAGAGGAGGCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1617
Qy 1305 CGATTTTCAAGGAGGATTAACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1364
Db 1618 CCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1677
Qy 1365 TGCTGATTAATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1424
Db 1678 TGCTGATTAATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1737
Qy 1425 ACTA---CAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1481
Db 1738 ACTATGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1797
Qy 1482 TTG 1484
Db 1798 TTG 1800

RESULT 3
AY081764 1695 bp mRNA linear PLN 31-MAY-2002
LOCUS Talaromyces emersonii beta-glucosidase (bg11) mRNA, complete cds.
DEFINITION
ACCESSION AY081764 GI:21284364
VERSION
KEYWORDS
SOURCE
ORGANISM
Talaromyces emersonii
Talaromyces emersonii
Bukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Talaromyces.
REFERENCE
AUTHORS 1 (bases 1 to 1695)
Murray, P.G., Collins, C.M. and Tuohy, M.G.
TITLE The molecular cloning of the beta-glucosidase genes of Talaromyces emersonii
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1695)
Murray, P.G., Collins, C.M. and Tuohy, M.G.
AUTHORS Direct Submission
TITLE Submitted (04-MAR-2002) Biochemistry, National University of Ireland, Galway, Ireland
JOURNAL 3 (bases 1 to 1695)
Murray, P.G., Collins, C.M. and Tuohy, M.G.
AUTHORS Direct Submission
TITLE Submitted (31-MAY-2002) Biochemistry, National University of Ireland, Galway, Ireland
JOURNAL Sequence update by submitter
REMARK
COMMENT On May 31, 2002 this sequence version replaced gi:19568099.
FEATURES
SOURCE
Location/Qualifiers
1..1695
/organism="Talaromyces emersonii"
/mol_type="mRNA"
/db_xref="taxon:68825"
1..1695
/gene="bg11"
107..1576
/gene="bg11"
/EC_number="3.2.1.21"
/note="family 1 glycosyl hydrolase"
/codon_start=1
/product="beta-glucosidase"
/protein_id="AI19551.2"
/db_xref="gi:21284365"
/translation="MTTETPTTTTSTLPPDLMGFATSYOIEGAFDEDDRLPSIT
DTEFSTPGKVEDGNDVACDSYHRTGEDIILKYGAKYRFLSPRIIPLEGND
PINKGQFYSKFDLHAAGIEPFTLFDLPELMKRYGGLNEEFVADANYA
RVNENALGSKYKWTITNEBPCSVLHGNDKAPGTSRTSSPEDGTEADYVH
NLVAHGTVVDIYRERFKOGEGITGNDMAEPMDPNADVEKCDKIEPAISM
PADPIYHKYDSDSVKQGLRPLKFTFBEILAFHGSNDFFGMNYCENTYLRNTRGEPD
PEDIAGLDLIMEDKSNPIGPEYQCEMLRPFPLGFRKLKMLADRYNNPKIYVTENG
TSVKGSDKPLEEVLNDEFRQYRYDIAMVAVADGVAVKAYVMAMSLDNEFWESE
GYRSRFGVTYVDYKNGQKRIPIKSAVLITGELFNKYRKE"

gene
CDS
ORIGIN

Query Match 22.7%; Score 452.2; DB 8; Length 1695;
Best Local Similarity 59.5%; Pred. No. 2e-81;
Matches 839; Conservative 1; Mismatches 559; Indels 12; Gaps 4;
QY 78 CTCTGCCCAAGATTGTAATGAGGCTTCGCAAGCGCGCTACAGATCGAAGCGCGC 137
Db 147 CTCTTCCCTCCGACCTTCCTCTGGGGTTTCGCACTGCAGTACCAATCGAAGGCGCT 206
QY 138 TCAAGAAGGTGGCGCGCGCCGCTCATCTGGGACAGTACTGCGACCTGGAGCCATCGC 197
Db 207 TCGACGAGGATGGCGCTCTCGATATGGGACACTTTCAGCAAGCGCGCGCAAGG 266
QY 198 GCACCAACGGCGCAAGCGGAGTGGCTTGCATCTACCAACCGCTACAGATGAGACT 257
Db 267 TCGAGGACGGGACCAATGGCGATGCTCCGCTGCACTGACCAACCGTACCGGTAGACA 326
QY 258 TTGATCTCTTGACCAAGTGGCGCAAGGCGCTACCGCTTCTCTGTCGTGGCGGGA 317
Db 327 TTGAGATATTGAAGAATAGCGGTGCCAAGATCTATCCCTTCTCTGCGCACGTA 386

QY 318 TCATTCCTTCGGGGGCGAGCTGATCCGCTCAACGAGAGGGAATTGATTTACAGCA 377
Db 387 TCATTCCTTCGGGGGCGAGCAAGCAAGCAATGATGAAAAGGGCGCTGACATTTACCA 446
QY 378 AACTGATTGAGCCCTTGTAGAGCGGGGTATCACGCTTGGGTGATCTTTGACACTGGG 437
Db 447 AGTTCTTGATGACTTACAGCGCGGGCATTCAGCCCTTCTGACGCTTTCCACTGGG 506
QY 438 ATTCCTTCAGGCGCTTACAGATCGCTATGAGAGCTGAGCTCAACGTTGAAGAGTCCAGC 497
Db 507 ATTCGCCGACGAGCTGATGAGCGGTACCGGGGCAATGCTCAACAGAGAGATTGTGG 566
QY 498 TGAATTTAGCGGATTCGAGAGGTGTGCTTTGACAGTTTGGGAGCCGATCCAGACT 557
Db 567 CCGACTAGCGCAACTATGCGCGGTGTGCTTCAACCGCGCTCGAAGTCAAGACT 626
QY 558 GATACACATCAACGAGCCCTGATTCAGGCACTATGATATGCAACCGGACAGCAAGC 617
Db 627 GATCACTTTCAGAGCCCTGTGCACTTCCGTCTCGGCGCAACACCGGACAGCAGC 686
QY 618 CCGCGGCGAGGAGAGCATTAACAGCACTCCAGGCGCAACCTGCCACTGACCGCT 677
Db 687 CCGCGGCGGAGAGAGCATCGAGCAAGAAAGTCCGAGGAGAGCGGCACTCGCAACTT 746
QY 678 GGCCTCGTGAAGAGCCGATATGATGAGCCCGCGCGCTGCTTACAGAGG 737
Db 747 GATCTGTGATCAACACTTCTGTGCTATGATACCTGTGAGCACTTATCGAGAG 806
QY 738 ACTTTCGCCCTTCGCAAGAGGCGAGATCGGCACTTCGCTCAACGCGCACTACTATAGC 797
Db 807 AGTTCAAGAGAAACAAGGCGGGGAATGGATATCAACCTCAACGCGGACTGGCGGAGC 866
QY 798 CCTGGAACAGCATATGAGCTTCGAGGAAAGAGAGCTCTGAGCGACGATGGAATTTCA 857
Db 867 CATGGATTCGGAACCAAGAGAGCTGAGACCTGCAACCGAAGATCGAATTCGCA 926
QY 858 TTGGCTGTTTGGCAATCCCATCTTCTTGAAGAGACTATCCAGAGAGATGAAGAG 917
Db 927 TCTGTGTGTTGGGAGACCGATTTACAGGGAATAC--CGGACAGATGTGTAAC 983
QY 918 AGCTGAGCGAGAGCTTTCAGCCCTCACTCCGCGGACTTTGGCATCTCAATCGCGAG 977
Db 984 AGTTGGGCGGACCTGTGCGAAGTTTCAACCGGAAATCGCTTCGAGCA--CGGCA 1040
QY 978 AGACGACTTCAAGGCAATTAATTAACAATCCAGTTCCGCGCGCACTTAAGAGGTC 1037
Db 1041 GCAAGATTTTACGAGTAAACACTACTGCGAAATTAACATCGCAACCGGACCGGCG 1100
QY 1038 CCGTCCCGAGAGAGACTATCTCGGCGCATCATGAGCAGCAGAGAAATTAAGAGCGCA 1097
Db 1101 AACCAAGCCGAGAGACATTTGCTGGAACATCTGATGAGAGAGCAAGAACGCA 1160
QY 1098 GCGCCGTTGGCGAGAGAGGCGCTGCTGCTGCGCTCTGCGCGCATGATTCGGA 1157
Db 1161 ACCCATTCGCGCCGAGCAACCAATGAGTGGCTGCGCGCTGTCGCGGTTTACAGA 1220
QY 1158 AGCATCTCGCGCGGCTGTACGAGCTGTAGCGAAGCC--ATTCATCATCAGCGAAGC 1214
Db 1221 AGCTGCTGAAGTGTCTTGAGACCGGTACCAACCCCAAGATCTACGTGACCGAAGC 1280
QY 1215 GATGCCGTCCTGAGAGAGAAATATGACGTGAGAGAGCGCTCAACGACCCCTTCC 1274
Db 1281 GAGCAGAGCTCAAGGCGAGAGAGCAAGCCCTCGAAGAGGTGCTCAACGATGATTC 1340
QY 1275 GCATCCGTACTTTGACTGCACTTGAGACTGATTTCCAGGCTATTCAGAGCGCG 1334
Db 1341 GCGTGAAGTACTACCGGAGCTATATCGGCGCATGTGATCTCTTGGACGAGATGGG 1400
QY 1335 TCGTGTCAAGGAGTACTTGTGAGGCGGTCTGATTAATCTGGAATGATGATGGCT 1394
Db 1401 TCAAGTCAAGGCTTACATGCTTGGAGATTGTGACAACTTCAAGATGCTCGAAGGCT 1460

RESULT 4	AB003110	2171 bp	DNA	linear	PLN 25-DEC-2002
LOCUS	Hypocrea jecorina bg12 gene for bete-glucosidase, complete cds.				
DEFINITION	AB003110				
ACCESSION	AB003110.1	GI:4249561			
VERSION					
KEYWORDS					
SOURCE	Hypocrea jecorina				
ORGANISM	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.				
REFERENCE	1 Takahashi,S., Nakamura,A., Hidaka,M., Masaki,H. and Uozumi,T. Molecular cloning and expression of the novel fungal bete-glucosidase genes from <i>Hemicola grisea</i> and <i>Trichoderma reesei</i> J. Biochem. (1999) In press 2 (bases 1 to 2171) Uozumi,T. Direct Submission Submitted (09-MAR-1997) Takeshi Uozumi, The University of Tokyo, Department of Biotechnology, Faculty of Agriculture; 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan (E-mail: uozumi@mbc.b.t.u-tokyo.ac.jp, Tel:03-5664-0387, Fax:03-5684-0387) Location/Qualifiers 1. .2171 /organism="Hypocrea jecorina" /mol_type="genomic DNA" /strain="QM9414" /db_xref="taxon:51453" /note="anamorph: Trichoderma reesei" join(556..596,670..2029) /gene="bg12" join(556..596,670..2029) /gene="bg12" /codon_start=1 /product="bete-glucosidase" /protein_id="BAI74959.1" /db_xref="GI:4249562" /translation="MLPKDFQWGFATAYOIEGAVDDQGRPSITWDFCAOPKTIAG SSGTTACDSYRTKREDIALLSLKANSRFSISRIIPBEGRDANYAGIDHYKVF VDLLDAGITPFTLFHWDLEGHJHORGGLINTPEPLDENARVWFRLPVRW ITFNEPLCSAIPGVSGFAPRGQSTSPFWGHNIIIVAHGKAVKAVADPFKPSGCG QIGVLNGDFPYPMWDADPADKKAERLLEFTWADPIYLGDPASMRKQLDRLE TFGEERLALVGSNDPFGMHNYSNYIRHRESPASADDTGVNVVLFPTNQGNCIGEE TOSWLRBPCAGPRDPLVWISKRYGPIPIYITENGSTIKESDLPRKILLEDPRVYK YNEIRAVTAVELDGVNVKGYFAMSLMDNFEMADGVTRFGVTYVDENGOKRFPKK SAKSLKPLDELIAAA"				
FEATURES					
SOURCE					
JOURNAL					
AUTHORS					
REFERENCE					
TITLE					
ABSTRACT					
KEYWORDS					
ORIGIN					
Query Match	20.0%	Score 397.6;	DB 8;	Length 2171;	
Best Local Similarity	59.4%	Pred. No.2.4e-70;			
Matches 824;	Conservative 1;	Mismatches 515;	Indels 48;	Gaps 7;	
Oy	106	CGCAACGGCCGCTACACAGATCGAAGCGCGCTCAAGAAAGATGGCCGCGCCGCTCAT	165		
Db	658	CCCGTCGACAGCTACAGATCGAAGCGCGCGCTGACACAGACGGCCGCGCCCAAGAT	717		
Oy	166	CTGGGACACGTACTGCCACCTGGAACCATCGCGACCAACGGCGCAACGGCGATGGCC	225		
Db	718	CTGGGACACGCTTCCGCGGACGCCCGCAAGATCGCCGACGCGCTGTCGGCGCTGACGCGC	777		
Oy	226	TTGGGATCACTACCAACCGCTACAGATGAGAGACCTTGATCTCTTGACCAAGTACGGCGCAA	285		
Db	778	GTGGACCTCGTACCAACCGCAGCGCCGAGAGACATTCGCGCTCAAGTCTCTGGGGCCAA	837		

OY	286	GGCCATACCGCTTCTCTTGATGAGTATCTCCCTCGCGGCAAGGCTGATCC	345
Db	838	GAGCTACCGCTTCTTCCATCTCTGTGTCGCGATCATCTCCCGAGGCGCGCGAGTGC	897
OY	346	CGTCAACGAGGAGGGAATTGATTTAACGCAAACTGATTGACGCTCTTTGAGGCGGAG	405
Db	898	CGTCAACGAGGCGGCGCATCGACCACTACGTCAAGTTGCTGCAGCAGCTGTCTGACGCCG	957
OY	406	TATACCGCTTGGGTGACTTTGTACCACTGGGATCTGCTCAGGGCGTTCAAGATCGCTA	465
Db	958	CATACGCTCTTCAATCACCTCTTCCATGTGGACTGTCCGAGGCGCTGCATCAAGCGGTA	1017
OY	466	TGAGAGCTGAGCTCAACGTGAAGAGAGTCCACTGGACTTTGAGCGGTTGCGAGGTTTG	525
Db	1018	CGGGGGCGTGTGAACCGCACCGAGTTCCGCTCGACTTTGAAAACTAACGCCCGCT---	1074
OY	526	CTTTGAACGTTTTGGGAGCCGAGTCCAGAACCTGATCAACATCAACGACCCTGATTTCA	585
Db	1075	CATGTTACGAGCGGCTGCCAAGAGTGCGCAATGTGATCTCTTCAACGAGCGCGTGTCTC	1133
OY	586	GGCATCTATGGATATGCAACGGGAGCAAGCCCCGGGAGAGGAGACATTTAACAAACA	645
Db	1135	GGCCATCCCGGACTACGGCTCCGGACTTGTGCCCCGGCGGAGACACA-----	1188
OY	646	CTCCACCGAGGGCAACACTGCACTGAGCCGTGCTCGCTGGAAAAGGCCAGATCATGAG	705
Db	1185	-----CTTGGAGCGTGGACCGTCCGCCACAAATCTCTGTGCG	1222
OY	706	.CCATGCCCGCCCGTGGCCGTCTACAGCAGGAGCTTTGCCCTTCGCAAA---AGGCCCA	762
Db	1225	CCACGCGCGCGCCCTCAAGCGGTACCGGAGCAGACTTCAAGGCCCGCAGCGCGCAGCGCA	1284
OY	763	GATGGAGACTCTCGCTCAACGGCGGACTACTATAGAGCCCTGGGACAGCAATGAGCCCTCGGGA	822
Db	1285	GATGGAGACTCTCTTCAAGGCGAGCTTACCTAACCTCTGGGAGCGCGCGCCAGCCGCGCA	1344
OY	823	CAAGAGGCTGCTGAGCGACGAGATGGAATTTACATTGGCTGTGTTTGCATTCCTCATTT	882
Db	1345	CAAGAGGCGGCGCGAGCGCGCCCTCGAGTTCTTACCGGCTGTGTTCCGGATATCCATCTA	1400
OY	883	CTTGAAGAAGACTATCTCAGAGAGCATGAAAGACGCTGGGCGAGAGCTTCCAGCCCT	942
Db	1405	CTTGGGC---GACTAACCGGCGGTGAGTGGCGCAAGCAGCTGGCGGCAACCGCTGCCGACTT	1463
OY	943	CACGCCGCGGACTTTGGCATCTCTCAATGGCGGAGAGCGACTTCTACGGGATGAATTA	1002
Db	1462	TACGCGCAGGAGCGCGCCTCTGTCCA---CGGCTCCAGCAGCTTTACGGGATGAAACA	1518
OY	1003	CTAGACATCCGAGTTGCGGCGCACTTGAACGCTGCCGTCCCGGAGACGCACTATCTCGG	1062
Db	1519	CTACACGTCAACTACATCTCGGCAACCGGAGCTCGCGCGCTCCGCGAGCAACCGTGG	1578
OY	1063	CGCCATCATAGACCAAGAGAAATTAAGACCGGACGCCGTTGGCGAGAGAGCGGCTT	1122
Db	1579	CAAGTCCACCTGTCTTTCACCAACAAGCAAGGCAAACTGCACTCGGCCCGGAGACGCACTC	1638
OY	1123	GCGCTGGTGGCTCTCGGCCGCGACATGTTCCGGAAAGCATCTCGGCCGAGGTATACGGCT	1182
Db	1639	CCCTTGTGCGCCCTCTGTGCGCGGCTTCCGGACTTCTGTGTTGATCAGCAAG	1692
OY	1183	GTACGGCA---GCCCATCTACATCACCGAAGACGAGTACCGCTCGAGAGAGAA	1233
Db	1699	GTAACGCTACCGGCCATCTAGTAGAGGAAGACGCGACGAGCATCAAGGGCGAGAGCGA	1758
OY	1240	CATGACGTGCGAGAGCGCGCTCAACGACCTTTCGCACTCCGATCTTGAATCTCGCACTT	1299
Db	1759	CTTGCCCAAGAGAGAGATTCTGAAAGATCTTCAAGGTCAAAGTAACTAATGAGTAT	1818
OY	1300	GGACTCGATTTTCAAGGCGATTACCCAGAGACGGCGTGTCTCAAGGGGTACTTTGCGTGG	1358
Db	1819	CGGTGCAATGTTACCGCGCTGAGACTGGAACGGGTCAACTCAAGGGGTACTTTGCTG	1878

QY	1360	GCGCTTCTCATTA	CTTGGAAATG	GTCAAGATG	CGTCAAGACCA	TTGGCGCTC	CAGCTT	1415
DB	1879	GTCCGCTATGGA <td>CAACTTTGAT <td>GTGGCGGCA <td>GGCGTCA <td>GTGA <td>CGATTTGGGGTTA <td>1938</td> </td></td></td></td></td>	CAACTTTGAT <td>GTGGCGGCA <td>GGCGTCA <td>GTGA <td>CGATTTGGGGTTA <td>1938</td> </td></td></td></td>	GTGGCGGCA <td>GGCGTCA <td>GTGA <td>CGATTTGGGGTTA <td>1938</td> </td></td></td>	GGCGTCA <td>GTGA <td>CGATTTGGGGTTA <td>1938</td> </td></td>	GTGA <td>CGATTTGGGGTTA <td>1938</td> </td>	CGATTTGGGGTTA <td>1938</td>	1938
QY	1420	CACAGACT---	ACACCA	CCCTCAAG	CGCACCGCCCA	AGAA	AGTCTGCTCTCA	1476
DB	1939	TGTGATTA	TATGAGAT	TGGGCA	GAGAGCGGTT	CCCCA	AGAAAGCGCA	1998
QY	1477	CATGTTTG	1484					
DB	1999	GCTGTTTG	2006					
RESULT 5	AY343988							
LOCUS	AY343988							
DEFINITION	AY343988	1590 bp	mRNA	linear	PLN 13-AUG-2003			
ACCESSION	AY343988							
VERSION	AY343988.1							
KEYWORDS	AY343988.1	GI:33521681						
SOURCE	Trichoderma viride							
ORGANISM	Trichoderma viride							
REFERENCE	Trichoderma viride							
AUTHORS	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;							
TITLE	Hypocreales; Hypocreales; mitosporic Hypocreales;							
JOURNAL	Trichoderma.							
REFERENCE	1 (bases 1 to 1590)							
AUTHORS	Liu, B. D., Yang, Q. and Zhou, Q.							
TITLE	Cloning and Sequence Analysis of the beta-glucosidase 2 Gene from							
JOURNAL	Trichoderma viride AS 3.3711							
REFERENCE	2 (bases 1 to 1590)							
AUTHORS	Liu, B. D., Yang, Q. and Zhou, Q.							
TITLE	Direct Submission							
JOURNAL	Submitted (15-Jul-2003) Department of Life Science and Engineering,							
FEATURES	Harbin Institute of Technology, Harbin, Heilongjiang 150001, P. R. China							
source	Location/Qualifiers							
gene	1..1590							
CDS	/organism="Trichoderma viride"							
	/mol_type="mRNA"							
	/strain="AS 3.3711"							
	/db_xref="taxon:5547"							
	1..1590							
	/gene="bg12"							
	192..1544							
	/gene="bg12"							
	/codon_start=1							
	/product="beta-glucosidase 2"							
	/protein_id="AA021384.1"							
	/db_xref="GI:33521682"							
	/translation="MLPNDPQWGFATYAAIYRGAVDQDGRGFSIMDTFCAPGKIADQ							
	SSGVADCSVYNRTAEDIALIKSLGAKSYFSSISWRIIPBEGRGAVNAQAGIDHYVKK							
	VNDLDAGITPPTLPEHMDPEGLHQRVGLNTEPEPLDFENYARVNFRLPKVRYM							
	ITFNEVIGSAIPYGSQSGTAPRGOSTSEBWTGHNILVAHGAVVAYDDPKASDDQ							
	OTGVNIGPTPIYPMADPADPADKEAERLEPFTAMPADPIYAGDYPAASMKQLGRRL							
	TTTPPERALVHSSNDYFQNMHTSYIRKRSRPAATTSATSTCSPTSRATASAPR							
	SPPLAVRPAATSWCGSARGITPIYVTENGSINGESDLPEKILIEDDFRKVINEI							
	IPAMVAVELDGVNVKGYPAWSIMDNFEWADGYVTRFGTVYVDYENGGKRSFRAQR							
ORIGIN								
Query Match	18.8%;	Score 373.4;	DB 8;	Length 1590;				
Best Local Similarity	59.4%;	Fred. No. 2e-65;						
Matches 802;	Conservative 1;	Mismatches 497;	Indels 51;	Gaps 8				
QY	81	TGCCCAACGACTTTGATGGGGCTTGCAACAGCGCGCCTTACAGATCGAAGGCGCGCTGA	140					
DB	196	TGCCCAAGAGACTTTGAGTGGGGTTGCCACGCGCTGCTTACAGATCGAAGGCGCGCTG	255					
QY	141	AAGAAGGTGCGCGCGCGCGCTTCATCTGGGACACGTACTGCCACCTGGAGCAATCGCGCA	200					
DB	256	ACGAGACGCGCGCGCGCGCGCGATCTGGGACACGTTTCCGCGGACGCGCGCGCAACATG	315					

OY	201	CCAAAGCGCCAAAGCGGATGTGGCTTCGATCACTACACACCGCTACGATGAGAACCTTG	260
Db	316	CCGACGGCTCCTCGGGGGGTGACCGCGGTGCACTCGTACAAACCGCACGGCGAGGACATGT	375
OY	261	ATCTCTTGACCAAGTACGGCGCAAAAGGACTACCGCTCTCTCTGTCTGTGTGCGGATCA	320
Db	376	CGCTGTCAAGT	435
OY	321	TTCCCTTCGCGGAGGCTGTGATCCCGTCAAACGAGGAGGAAATTAAGTTTACGCAAC	380
Db	436	TCCCGAGGGGCGGCGCGGCGATCCCGTCAACGAGGGGCGATCGACACTACGTCAAGT	495
OY	381	TGATTTAGAGCCCTTTTGAAGGCGGGGTTTCAAGCTTGTAGGATCTTTGACACTGGGATC	440
Db	496	TGTGTACAGACTGT	555
OY	441	TGCTTCAGGCGCTTCAAGATGCTATGAGAGGCTGTCAACGTGAAAGAGTCCAGCTGG	500
Db	556	TGCGGAGAGGCTGTGATCAGCGGTACGAGGGGGCTGTGAACCGCACGAGTCCCGCTCG	615
OY	501	ACTTTGAGCGGTATGCGAGGT	560
Db	616	ACTTTGAAATACGCGCGCGCTCATGTTCAGGCGCT--GCCAAGGTGTGCAACTGGA	672
OY	561	TCACCATCAAGAACCTTGATTCAGGCACTTATGAGATATGCCACCGGACGACACGCC	620
Db	673	TCACTTTAAGAGCGCGT	732
OY	621	CGGCGAGGACGACTTATCAAGCATCTCACCGAGGGCAACATGCCACTGAGCGGTGGC	680
Db	732	CCGCGCGGCGAGGCA-----CCGTGAGCGGTGGA	762
OY	681	TGCGTGGAAAGGCCCATGATCATGAGCCATGCGCGCGCGGTGCTATACGAGGAGCT	740
Db	763	CGGTGGCGCAACATCTCTGT	822
OY	741	T---TCGCGCCCTCGCAAAAGGCGAGATCGGCACTCTGTCTACAGCGGCACTATATGAC	797
Db	823	TCAAGCCGCGCAGCGCGCGACGCGGAGATCGGCACTGTCTTCAAGCGGCACTTCACTAAC	882
OY	799	CCTGGGACAGCAATGAGGCTGTGGGACAGAGGCGTGTGAGCGAGCGATGTGAAATTTACA	857
Db	883	CTTGGAGCGCGCGGACCCCGCGGACAAAGAGGCGGCGGAGCGGCTGTGAGTTCTTCA	942
OY	858	TTGGCTGTGTGCGCAATCCCATCTTCTTGAAGAAAGATATCAAGAGCATGAAAGAC	917
Db	943	CGGCTGT	999
OY	918	AGCTGGGCGAGAGCTTTCAGGCTCTCACTCCCGGCGACTTTGGCACTCTCAATCCCGGAG	977
Db	1000	AGCTGGGCGAGCGGCTGTGCGCACTTTTACGCGGAGAGAGCGGCGCTGTCA---CGCT	1056
OY	978	AGACGCACTTCTAAGCGCATGATTTCTACACATCCAGTTGGCGGCGACCTAGACGCTC	1037
Db	1057	CCAAAGACTTTTACGCGCATGAAACATACAGTCAATCATTCGCGACCGGACGAGCTGC	1116
OY	1038	CGTGTCCCGAGAGCGGACTTCTCGGCGCACTCATGAGCACAGAGAAATTAAGACGCGA	1097
Db	1117	CCGCTC---GACGACACCGGTGGCAAGTGTGAGTGTCTTCAACCAACGAGGAGGCA	1172
OY	1098	GCGCCGTTGGCGAGAGAGCGGCGCTGTGTGGCTGTGCTGCGCGCGGACATGTTCGGA	1157
Db	1173	ACTGATGTGGCGCGGAGAG--GCGATGTCTTGTGCGCTGTGTGCGCTGTGCGCGACTTC	1231
OY	1158	AGCATCTCGCCCGGAGTGTACGCGCTGTACGCGAAGCCCATCTTACATCCGAGAAAGGAT	1217
Db	1232	CTGT	1287
OY	1218	GCGCGCTGT	1277
Db	1288	CGAGCATCAAGGCGAGAGCGACTTGTGCGCAAGAGAAAGATTTCTGAAATGACTTTCAGG	1347

QY	1278	TCGGTACTTGTGACTGCGACTGATTCGATTCAGGCCATTACCCAGACGGCGCTCG	1337
DB	1348	TCAAGTACTTAACGAGTACATCCGTCGATCCGCCGCGAGCTGACGCGGGCTCA	1407
QY	1338	TTCTCAAGGGGTCTTGTGCGTGGCGCTGCTCGATTAACCTTGGAATGTGTCAATGGCTACG	1397
DB	1408	ACCTCAAGGGGTCTTGTGCGTGGCGCTGCTCGATTAACCTTGGAATGTGTCAATGGCTACG	1467
QY	1398	GACCCAGATTCGGCGCTGACGCTTACAGACTA	1428
DB	1468	TGACGAGCTTTGGGGTTCATGATGTGATTA	1498
RESULT 6			
AK119546			
LOCUS			
DEFINITION	AK119546	1828 bp	mRNA
ACCESSION	AK119546		
VERSION	AK119546.1		
KEYWORDS	FLI CDNA; oligo capping.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE			
AUTHORS	1		
	The Rice Full-length CDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length CDNA Project Team; Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohnuki, K., Shishiki, T., Foundation of Advancement of International Genome Sequencing & Analysis Group; Otsomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurotsaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Nakikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN; Kawai, J., Carinci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Otsu, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shitaki, T., Yoshino, M., and Hayashizaki, Y.		
TITLE	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice		
JOURNAL	Science	301 (5631)	376-379 (2003)
MEDLINE	22752273		
PUBMED	12869764		
REFERENCE	2		
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carinci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Horii, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurotsaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakahama, Y., Nakamura, M., Namiki, T., Narikawa, R., Nishikura, J., Nishii, M., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohnuki, K., Oka, M., Ooka, H., Otsu, N., Ota, Y., Otsomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shitaki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.		
TITLE	Collection, mapping, and annotation of 28k full-length cDNA clones from japonica rice		

QY	1278	TCGGTACTTGTGACTGCGACTGATTCGATTCAGGCCATTACCCAGACGGCGCTCG	1337
DB	1348	TCAAGTACTTAACGAGTACATCCGTCGATCCGCCGCGAGCTGACGCGGGCTCA	1407
QY	1338	TTCTCAAGGGGTCTTGTGCGTGGCGCTGCTCGATTAACCTTGGAATGTGTCAATGGCTACG	1397
DB	1408	ACCTCAAGGGGTCTTGTGCGTGGCGCTGCTCGATTAACCTTGGAATGTGTCAATGGCTACG	1467
QY	1398	GACCCAGATTCGGCGCTGACGCTTACAGACTA	1428
DB	1468	TGACGAGCTTTGGGGTTCATGATGTGATTA	1498
REFERENCE			
AUTHORS	Unpublished		
TITLE	3 (bases 1 to 1828)		
JOURNAL	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)		
COMMENT	This clone is one of the 32k full-length cDNA clones from japonica rice.		
	URL: http://cdna01.dna.affrc.go.jp/cDNA/		
	NIAS Rice Full-length CDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohnuki, K., Shishiki, T., Yamamoto, M., and Nakahama, Y.		
	FAIS Genome Sequencing & Analysis Group: Otsomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurotsaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nishikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.		
	Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carinci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Horii, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Otsu, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shitaki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.		
	Location/Qualifiers		
FEATURES	1. 1828		
source	/organism="Oryza sativa (japonica cultivar-group)"		
	/mol_type="mRNA"		
	/cultivar="Nipponbare"		
	/db_xref="taxon:39947"		
	/clone="001-207-A05"		
ORIGIN			
	Query Match 14.0%; Score 278.2; DB 8; Length 1828;		
	Best Local Similarity 54.7%; Pred. No. 4.2e-46;		
	Matches 785; Conservative 0; Mismatches 608; Indels 42; Gaps 10;		
QY	81	TTGCCAAGCACTTGAATGGGCTTGGCAAGCGCCCTACACGATGGAAGCGCGCTCA	140
DB	208	TTCCCGAGGGGTCGCTTCCGACCGCTCCCGCGACGATCGAGGGAGCTGTGA	267
QY	141	AAGAAGTGGCGCGCGCGCTTCATCTGGGACAGTACTGCCACCTGAGCCATCGGCA	200
DB	268	AGAGAGACGGAGAGGCGACACCATCTGGGACAGCTTGGCGCACCTTTGGAAGATCA	327
QY	201	CCAAAGCGCGCAAGCGGATGGCTTGGGATCTACCAACCGCTACGATGAGAGCTTGG	260
DB	328	CGACCTTCAAGCAATGCTGATGTTGCGATGATGATGATGATGATGATGATGATGATG	387
QY	261	ATTCTTTCACCAAGTACGGCGCAAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	320
DB	388	AGCTTCAAGCAATGCTGATGTTGCGATGATGATGATGATGATGATGATGATGATG	447
QY	321	TTCCCTCGCGGACAGCTGATCCGCTCAAGAGGAGGAATGATTTACAGCAAC	380
DB	448	ACCCAATGCTG-----TTGCTCAAGTCAATGCTGATGATGATGATGATGATGATG	501
QY	381	TGATTAGCCCTGTTGAGCGGGGATATCAAGCTTGGTGAATTTTACACATGGGATC	440

Db 502 TGAATGATGCACTTCTAGCAAAAGAAATTACGCCATATGTGACACTTACCATCTGGAC 561
QY 441 TGCTCAGGCGCTTACGATCGCTATGAGAGCTGGCTCAACGTGGAAGAGTCCAGCTGG 500
Db 562 TCCCCAGGCGCTTGAAGAACAGTACAAAGGCTGCTTGAACAGCA---GATAGTGAACG 618
QY 501 ACTTTAGCGGATATGCGAGGTTGTCTTTGAACGTTTGGGAGCCGAGTCCAGAACTTGA 560
Db 619 ATTTCCGGGGGTACGCGAGACGCTGCTTCAAGGAGTTCCGGGAGAGGATGAAGCACTTGA 678
QY 561 TCACCATCAACGAAACCTGATTCAGGCACTCTATGATATGACACGCGAGCAACGCC 620
Db 679 TCACGCTCAACGAGCCGACACGCTGGCATTCAGGCTACGAGCAGAGGCTCCAGGCC 738
QY 621 CGGAGCAGAGCAGC-----ATTAAACAAGCACTCCACGAGGCAACACTGCCACTGAGC 674
Db 739 CCGGCGCGCTGCTCGGTGCTGCTTCCACCTCTACCAAGGCGGCACTCCGGCAGCCAGC 798
QY 675 CGTGGCTGCTGGAAGAGCCCAATCATAGACCATGCCCGGCGCTGGCCGCTTACAGCA 734
Db 799 CTTACGCTGCTGCCACCACTTCTCTCGCCGCGCGCGCGCAGCATCTACAGGA 858
QY 735 GGGACTTTGCGCCCTCGCAAAAGGCGCAGATCGGCATCTCGCTCAACGCGCACTACTATG 794
Db 859 CAAATACAGGCGACGAGAACGCGGAGCTTGGGATAGCGTTGACGTGATGTGTTGG 918
QY 795 AGCCCTGGGACAGCATGAGCCTCGGACAAAGAGGCTGTGAGCGACGATGGAATTTC 854
Db 919 AGCGGATGTCCAAACCAACGATC---GACATCGAGGCGCGCAAGAGCAGGAGATTC 975
QY 855 ACATTTGGCTGTTTGGCCAAATCCATCTTCTTGAAGAAAGCATATCCAGAGAGCATGAAGA 914
Db 976 AGTAGAGATGTTGCTGATCCGTTCTTT---CGGGAATACCCGGCGAGCATGAGGG 1032
QY 915 AGCAGCTGGGCGAGAGGCTTCCAGGCTTCCGCGGAGATTTGCCATCTCAATGCGG 974
Db 1033 CGAGGCTTGGGGAGAGGCTGCGCAGAGTTCAAGCGGATGAGGCGCGCTGCTCAAG---G 1089
QY 975 GAGAGACCGACTTCTTACGGCATGAATTCTACACATCCCATGTTGCGGCGCACTAGACG 1034
Db 1090 GGGCGCTGATTTCTTCGGCAATTAACCACTACACCACTACTACACGAGGCAACAACA 1149
QY 1095 GTCCGCTCCCGAGA-----CGGACTATCTGGGCGCATTCATGAGCAGCAGAGAAAT-- 1087
Db 1150 CCAACATATAGGGGACATTTGCTCAACAACACTTGGCAGACACCGGACACGTCAGCTCC 1209
QY 1088 -----AAGACGCGAGCCCGCTTGGCGAGAGAGCGGCTGCTGGCTGGCTCTGCC 1142
Db 1210 CATTCAAGATGGGAAGCAATTGGAGATGAGGCAAAATTCGATATGGCTGTACATTTGTC 1269
QY 1143 CGGACATGTTCCGGAGATCTCGCCCGGCTGTAAGGCTGT---ACGGCAAGCCCATCT 1199
Db 1270 CCGGAGGATGAGAGCTGATGAATATGTCAAGAAAGTAACAACGCCACACATGTT 1329
QY 1200 ACATCACCGAAGACGATGCGGCTGCGAGAGAGAACATGACGTGCGAGAGGCGC 1259
Db 1330 ACATCACTGAAGAACGGATGATGACAGCAACACCGTTCACTTTCATCAAGAGAGCGCC 1389
QY 1260 TCACGAGCCCTTTCGCGATCCGTAATCTTTGACTCGCATCTGGAATTCGAAGGCCA 1319
Db 1390 TCACGAGACGCAAGAGATCAATATACCAATGATACCTCACCAATCGCTGGCTGTCCA 1449
QY 1320 TTACCCAGAGAGGCGCTGCTGCTCAAGGGGTACTTTGGCGGGGCTGTGCTGATACTTGG 1379
Db 1450 TCACGAGAGAGCGGCTGCGACGTACGTGAGTCTTGGCGGTCTGCTGCTGCAACTGGG 1509
QY 1380 AATGCTCAGATGCTCAGACCCAGATTCGGGCTCAAGTTCAACAAGTCA---CCACCC 1436
Db 1510 AATGGGCGGCGGATATCTCTGAGATTCGGGCTCACTTGTGAGCTTCAAGAGTAAC 1569
QY 1437 TCAAGCGACGCGCAAGAACTCTGCTGTCTCAAGACATGTTTGGGCGCC 1491
Db 1570 TCAAGAGATACCCCAAGAACTCGGTGCAAGGCTCTCTGGAAGACCTG 1624

RESULT 7
AY129294
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) beta-glucosidase mRNA,
complete cds.
AY129294
VERSION
AY129294.1 GI:22658439
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 2032)
REFERENCE
Akiyama,T., Ogasawara,R., Ketudat-Cairns,J.R., Swast,J. and Easen,A.
Predominant expression of beta-glucosidase in germinating rice
seeds
JOURNAL
2 (bases 1 to 2032)
REFERENCE
Akiyama,T., Ogasawara,R., Ketudat-Cairns,J.R., Swast,J. and Easen,A.
Direct Submission
Submitted (05-JUL-2002) Low Temperature Science, National Hokkaido
Agricultural Research Center, 1 Hitsujigaoka, Toyohira-ku, Sapporo,
Hokkaido 062-8555, Japan
LOCATION/Qualifiers
1..2032
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Yukihikari"
/db_xref="taxon:39947"
142..1707
/codon_start=1
/product="beta-glucosidase"
/protein_id="AA01354.1"
/db_xref="GI:22658440"
/translation="MGRIKSSSGRCSTABAEVAIVLVFVGAASSLRGCTIAQSGG
LTPGSPGEGVFCTASAYOYEGAVEDGQITMPFATPKRIDFSNADVAVDY
HRPEEDIQIADMGMDAYRPSIAMSRTYRQGVNOAGIDHAKLIDALLAGIOP
VTIYHMDLPALMDKXGMIDROIUDPAVAVCTCPREDRVKHTLLEPHVALO
GYDAGLPGRCSVLHLHYKANSRGTEPVVAHHPTIAAASITRTVKATONCO
LGIAFDVMEPEMSNTTIDLEAKRQEPOLGMPADFPFGDYPATRAVGRILP
TADBAVVKGLDFVGINHTYTYRTRNNNTIIIGTLNLTLDITGLVSLPFGKGP
DRANSILYIVPRGMSIMNYKERYNSPVTYTEMGMDSNPFISIKDALDSKRI
KYHNDVLTNLAASIKEDGCDVRGYFAWSLIDNEMWAGYSRFGELYVDYKMLKRYP
KNSVQWPKALKLT"

ORIGIN
Query Match 14.0%; Score 278.2; DB 8; Length 2032;
Best Local Similarity 54.7%; Pred. No. 4,2e-46;
Matches 785; Conservative 0; Mismatches 608; Indels 42; Gaps 10;
QY 81 TCCCAACAGACTTGAATGGGGCTTCCGCAACGGCGGCTTCCAGATGGAAGGCGCGCTCA 140
Db 290 TCCCGAGGGTTGCTTCCGACCGCTTCGCGGTACGATACGAGGAGCTGTGA 349
QY 141 AAGAAGTGGCGGCGCCGCTCATCTGGGACAGTACTGCGCACTGAGACCATGGCGCA 200
Db 350 AGGAGAGCGGAGAGGCGACATCTGGGACACGTTCCGCGCACCTTTGGAAAGATCA 409
QY 201 CCAACGGGCGCAACGGCGATGCGCTTGCATCACTACACCGCTAGATGAGACTTGG 260
Db 410 CGACTTCAGCAATGCTGATGTTGATGATGATACACCGCTTTCGAGAGAGATATAC 469
QY 261 ATCTTTGACCAAGTACGAGCCCAAGAGCTTACCGCTTCTGTGCGTGGCGGATCA 320
Db 470 AGCTCATGCGACATGCGAGTGGATGCTATCGGTTTCATAGCATGATGCAAGAACT 529
QY 321 TTCCCTCGGCGGAGCGCTGATCCCGTCAACGAGAGGGAATGATGTTTACGCAAAAC 380
Db 530 ACCCAATGCTG-----TTGCTCAAGTCAATCAAGCTGATGACCACTACAACAAGC 583


```

OY 381 TGAATGACGCTTGTGAGCGGGGATATACGCGCTTGGTACTTTGTACCTAGGATC 440
    |||||
DB 584 TGATGATGACTTCTGACAAAAGGATTCAGGCAATATGTGACACTTACACTGGGACC 643
    |||||
OY 441 TGGCTCAGGGCTTACAGATCGCTATGAGAGCTGGCTCAACGTGGAAGAGTCCAGCTGG 500
    |||||
DB 644 TCCCGCAGGCGCTTGAAGACAGATACAGGCTGGCTTGAACAGCA---CATAGTGACG 700
    |||||
OY 501 ACTTTGAGGGGATGAGAGGTGTGCTTGAAGCTTTTGGGAGCCGAGTCCAGAACTGGA 560
    |||||
DB 701 ATTTGGGGGATGAGGAGACGCTTCAAGGAGTTGGGAGAGGAGGTTAAGACACTGGA 760
    |||||
OY 561 TCACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGCAGCAACGCC 620
    |||||
DB 761 TCAGCGTCAACGAGCGGACACGATGGCCATTCAGGGCTTACGACGAGGGCTCCAGGCC 820
    |||||
OY 621 CCGGCGAGGAGCAGC-----ATTACAGACATTCACCGAGGGGACACATGCGCACTGAGC 674
    |||||
DB 821 CCGGCGGCTGCTCGTGTGCTCTCACCCTTACTGCAAGGCGGCACTCCGGAACCGAGC 880
    |||||
OY 675 CGTGGCTCGTGGAAGGCCCAATCATAGCCATGCGCGCGCGCTGGCGCTGACAGCA 734
    |||||
DB 881 CTTACGCTGCGCCACCACTTACTCTCGCCACGCGCGCGCGCAGCATCTACAGGA 940
    |||||
OY 735 GGGACTTTCGCGCCTCGCAAAAGGGCGAGATCGGCATCTGCTCAACGCGCACTACTATG 794
    |||||
DB 941 CAATAATCAAGGCGAGCGCAGACGGGAGCTTGGGATAGGCTTGCAGTGTATGTGTTGG 1000
    |||||
OY 795 AGCCCTTGGGACAGCAATGAGCTTCGGGACAGAGGCTGTGAGCGAGCGATGGAATTC 854
    |||||
DB 1001 AGCCGATGTCAACACCCAGCATC---GACATCGAGGGCGGCAAGAGGCCAGAGTTTC 1057
    |||||
OY 855 ACATTTGCTGGTTTGGCAATCCATCTCTTGAAGAGATATCCAGAGACATGAGA 914
    |||||
DB 1058 AGCTAGATGTTTGTGATCCGCTTCTTCTT---CGCGCATACCGCGCAAGTATGAGG 1114
    |||||
OY 915 AGCAGCTGGGCGAGAGGCTTCAAGCCCTCACTCCCGGAGCTTGTGCATCTCAATGCG 974
    |||||
DB 1115 CGAGGCTGGGAGAGGCTCCGAGGTTCAAGCGGATGAGCGCGCGCTGTCTAG---G 1171
    |||||
OY 975 GAGAGACCGACTTCTACGCGATGATTAATCAACATCCAGTTTGGCGCGCACTAGACG 1034
    |||||
DB 1172 GGGCGCTGATTTCTGTGCGATTAACACATACACACTTACTACAGAGCACAAACA 1231
    |||||
OY 1035 GTCCCGTCCCCGAGA-----CGGACTATCTCTGGGCGCATCATGAGACCAAGAGAT-- 1087
    |||||
DB 1232 CCAACATCATCGGAGCAATGCTCAACAAACCTTGGGACACCGGACCGTACGCTCC 1291
    |||||
OY 1088 -----AAGACGCGAGCGCGCTTGGCGAGAGAGCGGCTCGCTGCTGCTGCTGCC 1142
    |||||
DB 1292 CATTCAGGATGGGAGGCAATGAGATAGGCGCAATTGGATATGCTGTACATTTGTC 1351
    |||||
OY 1143 CGGACATGTTCCGGAAGCATCTGCGCGGCTGTACGCGCTGT---ACGGCAAGCCATCT 1199
    |||||
DB 1352 CCGGAGGATGAGAGGCTGATGAACTATGTCAAGGAAAGTACAAACAGCCCACTGTG 1411
    |||||
OY 1200 ACATTCACCGGAGAGCATGCTGCGCGCTGAGAGAGGAACTATGCTGCGAGAGAGCCG 1259
    |||||
DB 1412 ACATTCACGTAAGAAAGGAGATGAGCAACAAACCGCTTCAATTCATCAAGAGACGCC 1471
    |||||
OY 1260 TCACAGACCCCTTCCGATCCGATCTGTAAGTCTGCACTTGAATTCATTTCCAGAGCA 1319
    |||||
DB 1472 TCAGAGGAGAGAGAGATCAAAATCAACAATGACTCTACCAATCTGGCTGTCCA 1531
    |||||
OY 1320 TTAACCAAGAGCGCGCTGCTCAAGGGGTAATTTGGTGGGCGCTTGCATTAATCTGG 1379
    |||||
DB 1532 TCAGAGGAGAGCGGGTGGAGTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1591
    |||||
OY 1380 AATGCTCAATGCTACGAGACCGGATTCGGCGCTCAAGTTTCAAGATATACA---CAACC 1436
    |||||
DB 1592 AGTGGGCGGCGGATCTCTCGAGATTCGGGCTCTACTTCTGTGAGATCAAGAGATACC 1651
    |||||
OY 1437 TCAAGCGCAGCGCCAAAGATGCTGCTGCTGCTCAAGACATGTTTGGCGCGC 1491
    |||||

```

```

DB 1652 TCAGAGATACCCCAAGAACTCGGTGCTGAGGCTTCAAGCGCTCTGAAAGACCTG 1706
    |||||
RESULT 8
AF268911
LOCUS
DEFINITION
Aspergillus niger beta-glucosidase precursor (bg11) gene, complete
cde.
ACCESSION
AF268911
VERSION
AF268911.2
KEYWORDS
GT:8572535
SOURCE
Aspergillus niger
ORGANISM
Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE
1. (bases 1 to 1988)
Al,Y.-C. and Meng,F.-M.
Molecular cloning and sequencing of a beta-glucosidase gene from
Aspergillus niger Ams11
Unpublished
JOURNAL
2. (bases 1 to 1988)
Al,Y.-C. and Meng,F.-M.
Direct Submision
Submitted (17-MAY-2000) Department of Biochemistry, Zhongshan
University, 135 Xingang Road, Guangzhou, Guangdong 510275, P. R.
China
REFERENCE
3. (bases 1 to 1988)
Al,Y.-C. and Meng,F.-M.
Direct Submision
Submitted (19-JUN-2000) Department of Biochemistry, Zhongshan
University, 135 Xingang Road, Guangzhou, Guangdong 510275, P. R.
China
REMARK
Sequence update by submitter
On Jun 19, 2000 this sequence version replaced gi:8248474.
COMMENT
Location/Qualifiers
FEATURES
1. 1988
/organism="Aspergillus niger"
/mol_type="genomic DNA"
/strain="Ams11"
/db_xref="taxon:5061"
<496..>1858
/gene="bg11"
join(<496..536,595..>1858)
/gene="bg11"
/product="beta-glucosidase precursor"
join(496..536,595..1858)
/gene="bg11"
/codon_start=1
/product="beta-glucosidase precursor"
/protein_id="AAF74209.2"
/db_xref="GI:8572536"
/translation="MLPKDLQWGFAPAAYQIEGAVDQGRGSIWDTFCAQPKIADG
SSGVTAQDSVYRPAEDIALKSLGASYSFSSSRIPBGRGDAVNAQAGIDHYKVD
DLIDAGITPEITLFHMDLHORYGILNTEFLDEBNARVAFRALPKVRNNEPLC
SAIPGYGSPAFGRQSTSEPTWGHNIIVAGRAVAKAYADPKFASGDOQIIVNG
DPTVPMADADPADKLELRFETAMFADPIYLGQVPAEMRQIDRLPTFPEBALVAG
SNDPYGMNHTSYIRHRSPPASADTVGVNVLFTNKGNCIGPTEPSEALVCAAG
FRDLVWTKSRGSPPIVYENGSTIKGSDLENKILDEDFVAKYNEITRAMVTAIV
ELDGVNVRGVTVYDENGQKRPKRSASLXLPFLDLAAA"
sig_peptide
join(496..536,595)
mat_peptide
/gene="bg11"
596..1855
/product="beta-glucosidase"
ORIGIN
Query Match 12.7%; Score 253.2; DB 8; Length 1988;
Best Local Similarity 56.0%; Pred. No. 4.9e-41;
Matches 708; Conservative 1; Mismatches 474; Indels 81; Gaps 9;
OY 106 CGCAAGCGCGCTACAGATCGAAGCGCGCTCAAGAGATGCGCGCGCGCTCCAT 165
    |||||

```

Db 583 CCGTCCGACGACGATCAAGATGAGGGGCGCTGACGACGAGACGGCGGCGCCAGCAT 642
 Qy 166 CTGGGACAGCTACTGCGCACTGAGCCATGCGGCAACAAGCGGCCCAAGCGGATGCGC 225
 Db 643 CTGGGACAGCTACTGCGCGGAGCGGCAAGATGCGGAGCGGCTGTCGGGCGGAGCGC 702
 Qy 226 TTGGGATCACTACCAAGCGCTACGATGAGAGCTTGTATCTCTTGAACCAAGTACGCGCAA 285
 Db 703 GTGGGACTCGTACACCGGACGCGGAGAGACATGGCTGTCAAGTGTCTCGGGGCGCAA 762
 Qy 286 GAGCTACCGCTTCTCTTGT 345
 Db 763 GAGCTACCGCTTCT 816
 Qy 346 CCGTCAAGAGAGAGAGATGATGTTTACAGCAACGATGAGCGCCGCTGTGAGCGCGG 405
 Db 817 CGTAAACGAGCGGCGGATGACCACTAGCTAAGTGTGTGACGACGCTGTGAGCGCGG 876
 Qy 406 TATCAGCGCTTGGGTGACTTTGTACCACTGCGGATCTGCGGCTTCAAGATGCGTA 465
 Db 877 CATACCGCTTCT 927
 Qy 466 TGAAGCTGCTCAACGTTGAAGAGTTCAGCTGAGCTTGTGAGCGGATGCGAGGTTGTG 525
 Db 928 CGGGGGGCTGTGAACCGCACCGAGTTCCGCTGACTTTGAAAACTACGCGCGCGTCTAT 987
 Qy 526 CTTTGAAGCTTTTGGGAGCGGATGTCAGACTGATCCATCAACGAGCGGCTGTGATTC 585
 Db 988 GTT-----CAGGGCGCTGCGCAAGGTGTGCACTGAGAGCGGCGCTGTGCTC 1035
 Qy 586 GGCATCTATGATATGCGCACCGGAGCGACGCGCGGCGGAGAGCACTTAAACAGCA 645
 Db 1036 GGCATCTCTGCGGCTACGCGCTCGGAGGTTGCGCCCGCGCGGAGAGCA----- 1085
 Qy 646 CTCAACGAGGAGCAACTGCGCTGAGCGCTGCTGCTGTGAGAGCGCGAGATATGAG 705
 Db 1086 -----CCTCGAGACCGTGTGAGACCGTGTGCGCAACATCTCTGTGTC 1125
 Qy 706 CCATGCGCGCGCGCGCTGCTACAGAGGAGACT---TGCGCCCTGCGCAAAAGGCGCA 762
 Db 1126 CCAAGCGCGCGCGCTGCAAGCGGTACCGGAGAGCTTCAAGCGCGCGGAGCGCGCA 1185
 Qy 763 GATCGGCTCTGCTCAACGCGGACTACTATGAGCGCTGAGAGCAATGAGCTCGGGA 822
 Db 1186 GATCGGCTCTGCTCAACGCGGACTTCAACCTACCGCTGAGAGCGCGCGGAGCGCA 1245
 Qy 823 CAGAGAGCTGTGAGAGCGAGGAGTTCATCTGAGCTGTGTGTGTGTGTGTGTGTGTGT 882
 Db 1246 CAGAGAG-----CGCTGTGAGTCTTCAAGCGCTGTGTGTGTGTGTGTGTGTGTGT 1293
 Qy 883 CTTGAAGAGAGCTATCCAGAGAGATGAGAGAGAGCTGGGCGAGAGCTTCCAGCCCT 942
 Db 1294 CTTGGAGC---GACTACCGCGGCTGATGCGAGAGAGCTGGGCGAGCGGCTGCCACCTT 1350
 Qy 1351 TACGCGCGGAGAGCGCGCTGTGCTCA---CGCTCAACGACTTTTACGCGATGAACA 1407
 Db 943 CACTCCGCGGACTTTGCGCTCTGATCTGATGCGGAGAGAGCGACTTCTACGCGATGAATTA 1002
 Qy 1003 CTACCATCTCCGAGTTGGCGGCGGACCTAGAGCGGTCCGCTCCCGAGAGAGATATCTGG 1062
 Db 1408 CTACCGCTTCAACTATCTGCGGCGGAGCTGCGCGCTGTGCGGAGAGAGCGGCTGG 1467
 Qy 1063 CGCCATCTATGAGAGAGAGAGATTAAGAGAGAGAGCGCGCGCTGTGCGAGAGAGCGGCT 1122
 Db 1468 CAGCTTCAAGCTCTTCTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1527
 Qy 1123 CGCCTGTGCTGCTCTGCGCGGAGAGATGTTGCGAGAGAGATCTGCGCGGCTGTACGCGCT 1182
 Db 1528 CCCCTGTGCTGCGCGCTGTGCGCGGCTTCCGCACTTCTGTGTGAGAGAGAGAGAG 1587
 Qy 1183 GTACGG---CAAGCCCATCTATACACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239
 Db 1588 GTACGGCTTCCGCGCGCTATCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1647

Qy 1240 CATGAGTGGAGAGAGCGCTCAACGACCCCTTCCGATCCGTACTTGTACTGCACTT 1299
 Db 1648 CTTGCCAAGAGAGAGATTTCTGAGATGACTTCAAGGTAAGTACTATTAACGATACAT 1707
 Qy 1300 GGAAGTATTTCAAGGCGCTTACCAAGAGAGCGCGCTGTGTGTAAGGAGTACTTGGTGT 1359
 Db 1708 CCGTGCATGTTTACCGCGCTGTGAGGCTGAGCGGAGTCAAGTGAAGTTTGGGATACGTA 1767
 Qy 1360 GCGG 1363
 Db 1768 TGTG 1771
 RESULT 9
 LOCUS AK067001
 DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013092F09, full insert sequence.
 ACCESSION AK067001
 VERSION AK067001.1 GI:32977019
 KEYWORDS FLI cDNA; CAP trapper.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.
 REFERENCE
 AUTHORS 1
 The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yanagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Oono, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Mura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Iehi, Y., Itoh, M., Kagawa, I., Kondo, S., Komoto, H., Miyazaki, A., Oono, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.
 TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
 JOURNAL Science 301 (5631), 376-379 (2003)
 MEDLINE 22752273
 PUBMED 12869764
 REFERENCE 2 (bases 1 to 2045)
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanegaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hizumoto, K., Hirooka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Komoto, H., Kondo, M., Koyda, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Mura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nishikura, J., Nishikura, K., Nomura, K., Numasaki, R., Ohneda, E., Oono, M., Ohtsuki, K., Oka, M., Ooka, H., Oono, N., Ota, Y., Oono, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Kahira, S., Tanaka, I., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yanagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.
 TITLE Direct Submission

JOURNAL

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT

This clone is one of the 28k full-length cDNA clones from japonica rice.

URL: http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project_Team_Kikuchi_S_Sato_K_Nagata_T_Kawagashira_N_Doi_K_Kishimoto_N_Yazaki_J_Ishikawa_M_Yamada_H_Ooka_H_Hotta_I_Kojima_K_Namiki_T_Onoda_E_Yanagi_W_Suzuki_K_Ii_C_Ohtsuki_K_Shishiki_T_Yamamoto_M

FAIS Genome Sequencing & Analysis Group: Otonari, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Naitikawa, R., Nishikawa, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsumoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Heshizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawaji, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Otsu, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Kahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES

Location/Qualifiers

1..2045
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultiar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013092F09"

ORIGIN

Query Match 12.5%; Score 248.6; DB 8; Length 2045;
Best Local Similarity 52.3%; Pred. No. 4.2e-40;
Matches 730; Conservative 1; Mismatches 632; Indels 33; Gaps 7;

81 TGGCCACGACCTTGAATGGGCTTCGCAACGGCCGCTTACAGATGAAAGGCGCTCA 140
256 TCCCGAGAGACTTCATCTTCGCGACCGGCTCCGCGCTTATCAGTACGAAGCGCTGTA 315
141 AAGAAGTGGCGCGCGCGCTTCATCTGCGACAGTACGCTGCACTTCAGGCAAAAGTTG 375
316 ACGAAGCGCGAGGCGCGAGCATCTGGGACACGACGCTCACTTCAGGCAAAAGTTG 375
201 CCAACGCGCGCAACGCGAGTGGCTTTCGATCACTACACGCGCTAGATGAGGACTTTG 260
376 AAGATGTAGCAACGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 435
261 ATCTCTTGAACGAAGTGAAGGCGCAAGGCGCTTACCGCTTCTCTTCTGCTGCTGCAATCA 320
436 ACTTCTGACCGACATGAAATGAGGCGCTTCTCTTCTGCTGCTGCAATGAGAGATTC 495
321 TTCCCTCGCGCGAGCGCTGATCCGCTCAACGAGAGGAAATTTGATTTAAGCAAAAC 380
496 TGCAGAAATGAAACATCACTGAGAGAAATCAAAAGGAGGAGTTGCTTCAAAACGCC 555
381 TGAATTGACGCGCTGTTGAGGCGGCTATCAAGCTTGGGTGATTTGATCACTGAGATC 440
556 TGAATTGACGAGCTATCTAGAGGCTTGAAGCATTTGTCACATCTTCACTTTTGACA 615
441 TGCCTGAGCGCTTCAAGATCGCTATGAGGCTGCTCAACGTTGAAAGAGGCTCCAGCTG 500
616 CCCACAGGCTCTGAGGACAAATACCGGAGCTTCTCA---GTGAAAACATTTGTGAAGG 672

Oy 501 ACTTTGAGCGGATATGCGAGGTTGCTTTGAAAGTTTGGGAGCCGAGTCCAGAACTGGA 560
Db 673 ATTTCGTGAGACTATGCGAGAGTGTGCTTCCGTGAGTTCCGCGACCGGTTGAGTATGA 732
Oy 561 TCACATCAACGAGCCGATGATTCAGGCGATCTATGATATGTCACCGCGAGCAAGCC 620
Db 733 ACACGTTCAACGAGCCGATGATTTCTGCGCGCGCGGCTACGGCTCCGACCAAGGCC 792
Oy 621 CCGGCGAGAGC-----ACATTAAACAACCTCCACGAGGCAACACTGCGACTGAGC 674
Db 793 CCGGCGGCTGCTCGCGTATGCTTCCAAAGAAATGCGCCCGGAGCTCGGCAAGAGC 852
Oy 675 CCGGCTGCTGGAAGAGCCAGATATGAGCCATCCGCGCGGCTGAGCTTACAGCA 734
Db 853 CGTACGTCGCGCGGCAACACTGCTGCTGCGCCACCGAGCGCTTCCATCACGCC 912
Oy 735 GGGACTTTGCGCTTCGCAAAAGGCGCAATCGCATCTGCTCAACGGCGACTATATG 794
Db 913 AAAAGTACAGGCGAGCGAGAAAGGCGAGATCGGCATCACGCGAGTGTGCACTGTTG 972
Oy 795 AGCCCTGGAGACGATGAGCTCGGAGACAGAGCTGCTGAGCGACGATGGAATTTTC 854
Db 973 TGCCT--ACAGCGAGCGCGCGCGGACAGACGCTGTCAGGCGCGAGCTTCACTTCA 1029
Oy 855 ACATTGCTGCTTGGCAATCCATCTTCTTGAAGAGAGATTCAGAGAGATGAGAGA 914
Db 1030 TGACGAGATGTTGATGAGACCGATGTTGTT---CGGCGACTTACCGGCGACATGAGA 1086
Oy 915 AGCAGCTGGCGAGAGAGGCTTCCAGCCCTGACCTCGCGAGCTTGGCTTCATGCGG 974
Db 1087 AGTTGCTGAGTACCGGCTGCGCCCAATTCACAGCGAGAGTGGAGTTGTGA---GG 1143
Oy 975 GAGAGCGACTTTCACGCGATGAAATTAATCAATCCAGTTGCGG-----C 1022
Db 1144 GATCTTACGACTTCACTGAGCTCTCACTTACTTACCAACAACTAGGTTAAGGTTTAC 1203
Oy 1023 GCCACCTTACGCGCTGCTGCGCGAGAGAGAGAGCGGCTGCGCTGCTGCTGCC 1142
Db 1204 GACGCGCGAGCAAGCTTCAAAACCGGCTTACGCTTACCAAGATCTTCACTTACGCGC 1263
Oy 1083 AGAATTAAGACGCGAGCGCGGCTTGGGAGAGAGCGGCTGCGCTGCTGCTGCC 1142
Db 1264 ACCGCAACGCGGCTCGGATCGGTCGCGGCTTTTCAAGATCTTCACTTACGCGC 1323
Oy 1143 CGGACATGTTCCGGAAGATTCGCGCGGCTGATGCGGCTTACGCGCAAGCCATCTTACA 1202
Db 1324 CGGCTGCTGAGTGTGCTGTTATACCAAGGAGATGACATGACATGATCTTACA 1383
Oy 1203 TCACGAGAAAGGATGCGCGTGCCTGAGAGAGAACTAGACGAGAGAGCGCTCA 1262
Db 1384 TTGCAAGAAAGGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1443
Oy 1263 ACAGCCCTTTCGCGATCGGATCTTGAATTTGACTCGCATCTGATTTTCAAGGCAATTA 1322
Db 1444 AGGATGACAAAGGATATCTTCACTTACAGATCTCGCT---TCATGACGCTGCA 1500
Oy 1323 CCGAGAGCGGCTGCTGCTCAAGGAGTACTTGGCTGCGGCTTCTGCTGATTAATTGGAAT 1382
Db 1501 TCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
Oy 1383 GGTGAGATGATGAGACCGAGATTTGGGCTGAGTTCACAGCTACACACCTCAAGC 1442
Db 1561 GGGGCGAGCGCTTACACCGCGCTTGGGCTCTTACGTCGACCGGAAATCTTCAAGA 1620
Oy 1443 GCACGCGCAAGAGT 1458
Db 1621 GGTATCGCAAGAGT 1636

RESULT 10
AF439322
LOCUS AF439322 2068 bp DNA linear JUN-2002

DEFINITION Talaromyces emersonii beta-glucosidase 1 (bg1) gene, complete cds.
ACCESSION AF439322
VERSION AF439322.3 GI:21493049
KEYWORDS
SOURCE Talaromyces emersonii
ORGANISM Talaromyces emersonii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Talaromyces.
REFERENCE 1 (bases 1 to 2068)
Murray, P.G., Collins, C.M. and Tuohy, M.G.
Molecular cloning and expression analysis of beta glucosidase from
Talaromyces emersonii
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2068)
Murray, P.G., Collins, C.M. and Tuohy, M.G.
Direct Submission
TITLE Submitted (23-OCT-2001) Biochemistry, National University of
Ireland, Galway, Ireland
REFERENCE 3 (bases 1 to 2068)
Murray, P.G., Collins, C.M. and Tuohy, M.G.
Direct Submission
TITLE Submitted (31-MAY-2002) Biochemistry, National University of
Ireland, Galway, Ireland
REMARK
REMARK 4 (bases 1 to 2068)
Murray, P.G., Collins, C.M. and Tuohy, M.G.
Direct Submission
TITLE Submitted (20-JUN-2002) Biochemistry, National University of
Ireland, Galway, Ireland
REMARK
REMARK Sequence update by submitter
On Jun 20, 2002 this sequence version replaced gi:21284362.
REMARK
REMARK Location/Qualifiers
FEATURES
source
1..2068
/organism="Talaromyces emersonii"
/mol_type="genomic DNA"
/db_xref="taxon:68825"
<304..>2068
/gene="bg1"
join(304..383,472..663,732..1204,1258..1843,1930..>2068)
/gene="bg1"
/product="beta-glucosidase 1"
join(304..383,472..663,732..1204,1258..1843,1930..2068)
/gene="bg1"
/EC_number="3.2.1.21"
/note="glycosyl hydrolase"
/codon_start=1
/product="beta-glucosidase 1"
/protein_id="AAL34084.2"
/db_xref="gi:21284363"
/translation="MTTETPTTTTSTLPDFLWGFATASQIEGAFDEGRPSIM
DPSKTPGVKVDGNTGVAACDSYHRTGEDIILKKYAKIRFSLSPRIILPLGRND
PINEKGLDYSKFLDDIHAAGIEPVTLPHNDLPDELMKRYGSMNKEPFADYANVA
RVFNALDSKVHMTTFNBPWCSSVLTNGHAPRTSDRKSREGSTPEMTVGH
NLVAHGTVDIYRREFKEKQGLGILTNGWAPWDPENADVEACDRKI ERIISW
PADPLVHGKVPDSWVKQGLDRLEKTFPEIARFVHNSNPFYGNHVCENYINRKEPD
PEDIIGNLDIIMEDKXGNPILGEPCEMLRPEPLGRKLKMLARVNNPKYVWENG
TSVKGESDPLIEVLENDERVOYRDYIGAWDAVOGVNKAVMASLIDNFEWSE
GTRSRGTVYVDYVKNQGRIPKSAVLVIGELFNKXIRKE"
ORIGIN
Query Match 12.4%; Score 247.8; DB 8; Length 2068;
Best Local Similarity 53.2%; Pred. No. 6,2e-40;
Matches 738; Conservative 1; Mismatches 518; Indels 130; Gaps 5;
QY 118 CTACCAAGATCGAAGGCGCCGTCAAAGAAGTGGCCGCGCCGTCCATCTGGGACACGTA 177
DB 472 CTACCAAGATCGAAGGCGCCGTCAAAGAAGTGGCCGCGCCGTCCATCTGGGACACGTA 531
QY 178 CTGGCCACTGGAGCACTCGCGACCAACGCGCCCAACGCGCTGTGGATCATCTA 237
DB 532 CAGCAAGACCCCGCGCAAGGTCAAGACGCGGACCAATGGCGATGTGGCTCGACATCGTA 591
QY 238 CCAACGCTACGATGAGGACTTTGATCTTTGACCAAGTACGCGCAAGGCGCTACCGCTT 297

DB 592 CCAACGCTACCGGTGAGGACATTGAGATATTGAAGAAATACGGTGCAGAGATCTATCGCTT 651
QY 298 CTCCTTTGCTGTGTG----- 313
DB 652 CTCGTGTCTCTGTGTGTCTCTGTGTATGTCTGCTCTTTATTTTTGTCTGTCTGTTTTT 711
QY 314 -----CGATTCATTTCCCTCTCGCGGAGCGTGCATCCCGTC 349
DB 712 TGTCTGACAGAAACAATCCAGGCGCATATCCTCTCTCGGGCGGAACGACCCCATTC 771
QY 350 AACGAGAGGGAATTGATTTTACAGCAATCTGATTAGCCCTGTGAGGGGGTATTC 409
DB 772 AATGAAAGGCGCTGAGTTCTACTCCAAAGTTCCTGATGACTTACAGCCGCGGCGATC 831
QY 410 AGCCCTTGGGTACTTTTACACATGGGATCTGCTCAGGCGCTTCAAGATGCTATGGA 469
DB 832 GAGCCCTTGTGTCAGCTCTTCCACTGGGATCTGCCGAGACTGATGAAGGGATACGG 891
QY 470 GGCTGCTCAACGTGAGAGGTCCAGCTGACTTTGACGGGTATTCGAGTTGTCTT 529
DB 892 GGCATCTCAACAGGAGGAGTTGTGGCCGACTAGCCAACTATGCCCCGTGTCTTTC 951
QY 530 GAACGTTTGGGACCGAGTCCAGAACTGGATCACATGAAGGCCGTGATTCAGGCC 589
DB 952 AAGCGCTCGGCTCGAAGGTCAAGCACTGATCTATTCAAGAGCCCTGTGTCACTTC 1011
QY 590 ATCTATGATATGCAACCGGACAGCAACGCCCGGCGAGAGCAGCATTAACAACACTCC 649
DB 1012 GTCTGCGCCACCAACACCGGACAGCACGCCCGGCGGACAGAGCATGAGCAGAAAGT 1071
QY 650 ACCGAGGCAACACTGCCACTGAGCCGTGCTGTGAAAGGCCCAATCATGAGCCT 709
DB 1072 CCGGAGGAAACGCACTCGCGAACCCTTGATGTGTGTACCAACTTCTGTGTGTCTCAT 1131
QY 710 GCCCGCGCTGCGCGCTCTACAGCAGGACTTTGGCCCTCGCAAAAGGCCAGATCGGC 769
DB 1132 GGTACGCTGTGACATCTATCGAAGAGCTTCAAGAGAAACAAGCGGGGAAATCGGA 1191
QY 770 ATCTGCTCAAC----- 781
DB 1192 ATCACTCTCAACGCTAGAGCTCGCGCTCATCTGATTCGATCTTATGTCTTAACCGCATCT 1251
QY 782 -----GGCGACTATATGAGCCCTGGGACAGCAATGAGCTCGGACAGAGAGCTGCTG 836
DB 1252 ATCTAGCGGACTGGCGCGGAACATGAGATCCGAAACCAAGACAGTCCAGAGCTGCG 1311
QY 837 AGCGACGATGGAATTTCAATTGGCTGTTTGCATCCATCTTCTTGAAGAAGACT 896
DB 1312 ACCGAAAGATCGAATTCGCATCTCGTGTTCGCGAGACCGGATTC---CAGGGGAAT 1368
QY 897 ATCCAGAGAGCATGAAGAAGACAGCTGGGCGAGAGGCTTCAAGCCCTCACTCCCGGACT 956
DB 1369 ACCGGAAGATCGAATTCGCATCTCGTGTTCGCGAGACCGGATTCACCCGGAAGAA 1428
QY 957 TTGCAATCTCTAATGCCGAGAGACCGAATCTTACCGCATGAATTAACAACATCCAGT 1016
DB 1429 TGGCTTCTGTGACAGGAGC---AAGATTTTATGCGATGAAACAATACTGCGAAAAT 1485
QY 1017 TGGCGCGCACTGACAGGTCCCGTCCCGAGACGGAATATCTTGGCGCCATCCATGAGC 1076
DB 1486 ACATTCGCAACCGGACCGGCAACCAACCGGAGACATTTGCTGGCAACTGGAATTC 1545
QY 1077 ACCGAGAGATTAAGACCGGACCGCCGTTGGCGAGAGAGAGGCGCTGCTGTGGCT 1136
DB 1546 TGATGAGAGACAAAGAACGCAACCCCAATCGGCCCGGAGACCAATGCGAGTGGTGGGC 1605
QY 1137 CTGCGCGGCAATGTTCCGGAAGCATCTCGCCGGGTGTACGGCTGTACGCGCAAGCCC- 1195
DB 1606 GCTTCCGCTGAGGTTCAGAGAGCTGTAAGTGTCTTGACAGACCGGTAAACAACCCCA 1665
QY 1196 --ATTAATCATACCGAAGCGATGCGCTGCGTCCGTGAGAGAGAACTGACGTGCGAGG 1253

Db 1666 AGATCTAGCTGACCGAGAACGAGCGCTCAAGGGCGAGAGCGACACCGCTCGAGG 1725
Oy 1254 AGGCGCTCAACGACCCCTTCGCGATCGGTAATCTTGACTGCACTTGATGATCTTCA 1313
Db 1726 AGGTGCTCAACGATGATGCTCGGCTGCACTACCGGACTATATCGCGCGATGTTGG 1785
Oy 1314 AGGCGATTAACGAGACGCGCTGCTCAAGGGGTAATTTGGCGTGGCGCTTGTCTGATA 1373
Db 1786 ATGCTGTGCAAGATGAGGCTCAAGCTTAAGCTTATGATGATGATTTGTTAAGT 1845
Oy 1374 ACTTGA 1380
Db 1846 AAGTGA 1852

RESULT 11
AX654334 1404 bp DNA linear PAT 22-MAR-2003
LOCUS Sequence 4204 from Patent WO03000898.
DEFINITION AX654334
ACCESSION AX654334
VERSION AX654334.1 GI:29157148
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Bristlegrass; Oryzaceae; Oryza.
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 1404
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN

Query Match 12.4%; Score 246; DB 6; Length 1404;
Best Local Similarity 54.5%; Pred. No. 1.5e-39;
Matches 725; Conservative 0; Mismatches 565; Indels 40; Gaps 10;

Oy 123 AGATCGAAGCGCGCTCAAGAGAGTGGCGCGCGCTTCATCTGAGACATCACTGCC 182
Db 86 AGTACGAGGAGCTGTGAAGAGAGAGCGAGAGGCGACCATCTGGGACACGTTGCGCG 145
Oy 183 ACCTGAGCCATCGCGACCAACGCGCAACGCGCATGTGGCTTGGATCACTACACC 242
Db 146 ACACCTTTGGAAGATCAACGACTTTCAGCAATGCTGATTTGCAATTGATCACTACACC 205
Oy 243 GCTAGATGAGACTTTGATCTCTTGAACAAGTACGCGCAAGCGCTTACCGCTTCTCT 302
Db 206 GTTTCAGAGGATATATACATCATGACACATGAGGATGATGCTATCGGTTCTGA 265
Oy 303 TGTGCGGTGCGCGATCATTCCTCGCGCGAGCGCTGATCCCTGTAAGAAGAGGAA 362
Db 266 TAGCATGTGTAAGATCTACCAATGTG-----TTGATCAAGTAAATCAAGCTGTGA 319
Oy 363 TTGAGTTTACAGCAAACTGATGACCGCTGTTGAGCGGAGTATCAACGCTTGGGTGA 422
Db 320 TGACACCACTACAAAGCTGATGATGCTTCTAGCAAAAGAAATTCACACCATATGTGA 379
Oy 423 CTTTGTACCACTGAGATCTGCTTCAAGGCTTCAAGTCTGATGAGGCTGCTCAAG 482
Db 380 CACTTACCACTGAGGACTTCCCGAGGCTTGAAGACAAGTACAGGCGCTGCTTGAACA 439
Oy 483 TGAAGAGGTCCAGCTGAGCTTTGAGGCGTATGAGAGTGTCTTTGAAGCTTTGGGG 542
Db 440 GG--CAGATAGTGAAGATTTGCGGCGTACGCGGAGAGCTGTTCAAGGAGTTCGGGG 496
Oy 543 ACCGAGTCCAGAACTGATGATCAACAGACCCCTGATTCAGGCGCATCTATGATATG 602

Db 497 ACAGGCTGAAGCACTGATATCAAGCTCAAGAGCGCACAGTGGGCATTCAGAGGCTACG 556
Oy 603 CCACCGGAGCAACGCGCCCGGAGAGAGCACTTA-----ACAAGACTCCACGAGG 656
Db 557 ACCGAGGCTCCAGGCGCCCGGCGCTGCTCGTGTCTCTCCACTCTATGGAAGCGG 616
Oy 657 GCAACATGCGCACTGAGCGGTGCTGTGAAGAGCCAGATCATAGAGCATGCCGCG 716
Db 617 GCAACTTCGAGCACAGCCCTACGTGTGTCGCCCAACATTCATCTTCGCCACGCGCGG 676
Oy 717 CCGTGGCCGTTCACAGAGGACTTTGCGCCCTCGAAGAGGCGAGTGCATCTGCG 776
Db 677 CCGCCAGCATCTACAGGACAAATTAAGGCGACGAGAAAGGCGAGCTTGGGATAGCGT 736
Oy 777 TCAAGCGGCACTATCTATGAGCCCTGGGACAGCAATAGCTCGGGAACAAGAGCTGCTG 836
Db 737 TGACCTGATGTGTGCTTGAAGCCGATGTCCAAACACAGAT--CGACATCGAGGCGGCA 793
Oy 837 AGGAGCGATGGAATTTACATTTGCTGTTTGGCAATCCATCTTCTTGAAGAGACT 896
Db 794 AGAGAGCGAGAGCTTTCAGCTAGATGTGTTGCTGATCCGTTCTT---CGGCGACT 850
Oy 897 ATTCAGAGAGCATGAAGAAAGACAGCTGAGGCTTCAAGCTTCACTCCGCGACT 956
Db 851 ACCGCGACAGATGAGAGGCGAGGCTGAGGAGAGGCTGCCAGAGTTCAAGCGGATGAGG 910
Oy 957 TTGCGATCTCTCAATGCGGAGAGACCGACTTTCACGCGATGAAATTAACATCCCACT 1016
Db 911 CCGCGCTGTGCAAG--GGGCGCTGTGATTTCTGTGCAATTAACCTACACCACTACT 967
Oy 1017 TCGCGGCGCACTTACAGGCGTCCGCTCCCGAGA-----CGACTATCTCGGCGCATCCA 1071
Db 968 ACACGAGGACACAAACACCAACATATCGGACATTTGCTCAACAAACCTTGGCAGACA 1027
Oy 1072 TGAACACGAGAGAAAT-----AAGACGCGAGCGCCGTTGGCGAGAGAGCGGCTCG 1124
Db 1028 CCGGACCGCTCAGCGCTCCCATTCAGAAATGGGAAGCAATTTGAGATTAAGGCAATTCGA 1087
Oy 1125 CCGTGTGCGCTCTCTCCCGGAGATTTCCGGAAGATCTCGCCGCGGTGTAAGGCTCT 1184
Db 1088 TATGCTGTGATCATTTGCGCCCGAGGATAGAGGCTGTGATGAACTATGTCAAGAAAGT 1147
Oy 1185 ---ACGGAAGCCCATCTATCATACCGAGAACGAGTCCCGTCCCTGGAGAGAGACA 1241
Db 1148 ACAACAGCCACCACTGATATCATGAAAGAGATCAAAATVCAAAATGATCACTCA 1207
Oy 1242 TGAAGTGCAGAGAGCGCTTCAACGACCCCTTCGCAATCCGTAATTTGAATCGCACTTGG 1301
Db 1208 TTTTCATCAAGAGAGCGCTTCAAGAGACAGAAAGATCAAAATVCAAAATGATCACTCA 1267
Oy 1302 ACTCGATTTTCAAGGCGCATTTACCAAGAGGCGCTGCTGTCAGAGGCTACTTTGCGTGG 1361
Db 1268 CCAATCTGTGCTTTCATCATAGAGAGAGGAGGCTACATGCTGTGATCTTGTGCGTGGT 1327
Oy 1362 CGTTGCTGCACTTCTTGAAG--TGTGAGATGCTTACGAGCCGATTTCCGCGTCAAGTTTC 1420
Db 1328 CTCTGCTGCACTGGAATTTGGGCGGCGGATATCTCTCGAGATTTGGGCTTCTACTTC 1387
Oy 1421 ACAGACTTACA 1430
Db 1388 GTGAGACTTACA 1397

RESULT 12
AK067231 2148 bp mRNA linear PLN 24-JUL-2003
LOCUS AK067231
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:013095012, full
insert sequence.
ACCESSION AK067231
VERSION AK067231.1 GI:32977249
KEYWORDS
SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1 The Rice Full-length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-length cDNA Project Team,
Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K.,
Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., Hotta I.,
Kojima K., Namiki T., Ohneda E., Yahagi W., Suzuki K., Li C.,
Ohtsuki K., Shishiki T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ootomo Y., Murakami K.,
Iida Y., Sugano S., Fujimura T., Suzuki Y., Tsunoda Y.,
Kurosaki T., Kodama T., Nasuda H., Kobayashi M., Xie Q., Lu M.,
Narikawa R., Sugiyama A., Mizuno K., Yokomizo S., Ninkura J.,
Ikedo R., Ishibiki J., Kawamata M., Yoshimura A., Miura J.,
Kusumegi T., Oka M., Ryu R., Ueda M., Matsubara K., RIKEN;
Kawai J., Carninci P., Adachi J., Aizawa K., Arakawa T., Fukuda S.,
Hara A., Hashidume W., Hayatsu N., Imotani K., Ishii Y., Itoh M.,
Kagawa I., Kondo S., Kono H., Miyazaki A., Osato N., Ota Y.,
Sato R., Sasaki D., Sato K., Shibata K., Shinagawa A., Shiraki T.,
Yoshino M., and Hayashizaki Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
Science 301 (5631), 376-379 (2003)

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

2 (bases 1 to 2148)

12869764

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T.,
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kawagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Miuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, K.,
Namiki, T., Narikawa, R., Ninkura, J., Nishii, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and
Yoshimura, A.

Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
Agricultural Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT
This clone is one of the 28K full-length cDNA clones from japonica
rice. URL: <http://cdna01.dna.affrc.go.jp/cDNA/>

NIS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K.,
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and
Yamamoto, M.
FALS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
Mizuno, K., Narikawa, R., Ninkura, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashizaki, K., Hayatsu, N., Hiramoto, K.,

Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.
Location/Qualifiers

FEATURES

source
1. organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013095012"

ORIGIN

Query Match 11.1%; Score 221.6; DB 8; Length 2148;
Best Local Similarity 55.9%; Pred. No. 1,3e-34;
Matches 529; Conservative 0; Mismatches 399; Indels 18; Gaps 5;

81 TGCCCAACGACTTGAATGGGCTTCGCAACGGCCGCTTACAGATGAAAGCGCGCTCA 140
Db TCCCGAGAGACTTCATCTTCGACACGGCTCCGCGCTTATCATAGAGGCGCTGTGA 262
Qy 141 AAGAGGTGGCCGCGCGCGCTTCATCTGGGACACGATCCGACCGAGCCATCCGCCA 200
Db 263 ACGAAGCGCGCAAGGGCCGAGATCTGGGACACGATCCGATCCATCCAGGCAACTTG 322
Qy 201 CCAAGCGCGCAACGGGATGTGGCTTGCATCTACACCGCTACGATGAGACCTTGG 260
Db 323 AAGATGTACCAACGGGATGTGACGATGACCTTTTACATCCGCTCAAGAGAGATCGA 382
Qy 261 ATCTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTGTGCTGTCCGATCA 320
Db 383 ACTTCGTGACCGACATGACATGACGCTTCGTTCTCATTTGAGGACGAGATCC 442
Qy 321 TTCCTTCGCGCGGAGGCTGATCCGCTCAACGAGGAGGAATGATTTACAGCAAC 380
Db 443 TGCCTAAATGAACCATCATGATGATGAAGAAATCAACAAAGAGGATTCCTTTTCAACAGCC 502
Qy 381 TGATTTGACGCTTGTAGGCGGGATATCAAGCTTGTGGTGTACTTTTACCATCTGGATC 440
Db 503 TGATTTACGAGTATATCTAGAGGTTGAGGATTTGACCATTTGTCATCTTCATTCGACA 562
Qy 441 TGCTTCAGGCGCTTACGATGCTATGAGGCTGGCTTCAACGTTGAGAGAGTCCAGCTGG 500
Db 563 CCCACAGGCTCTGAGAGACAAATACCGAGCTTCCTCA--GTGAATAATGTGTAGG 619
Qy 501 ACTTTGACGGTATGCGAGGTGTGCTTTGAACGTTTGGGGACCGAGTCCAGAACTGGA 560
Db 620 ATTTGTGACTATGAGGACGTGTGCTTCCGTGATTCGCGACCGGCTGAAATGATGGA 679
Qy 561 TCACCATCAACGAAACCTGTGATTCAGGCATCTATGATATGCCACGGACGAAACGCC 620
Db 680 ACAAGTTCAAGACCCATGATCTTGTGCGCGGCGGCTACGCTCCGACCAAGGCC 739
Qy 621 CGGCGAGAGC-----AGATTAAACAGACTCCACCGAGGGAACATGCTCCATGAGC 674
Db 740 CGGCGCGTGTCTCCCGTATCTTCCAAAGAAATCGCCCCCGGAGCTCGGCAACAGC 799
Qy 675 CGTGGCTGCTGGAAGGCCAGATCATGAGCCATGCGCGCGCGCTGAGCTTACAGCA 734
Db 800 CGTAGTGGCGGCAACAATCTGTCTGCTGCCACGCGGAGGCGCTCCGCTTACCGCC 859
Qy 735 GGGAGCTTTCGCGCTTCGCAAAAGGCGCAGATGCGCATCTGCTCAACGGCCATCTATG 794
Db 860 AAAAGTACACGACGACAGAGAGGCGCAGATCGGACATCAGCAGGTGTGCACTGTTGG 919
Qy 795 AGCCCTGGACAGCAATGAGCTTCGGGACAAAGAGGCTGCTGACGACGATGGAATTC 854

Db 920 TGCCCT---ACAGCAGCCGCGCCGACAGCAGCCGCTCAGGCGGAGCCCTCGACTTCA 976

Qy 855 ACATGGCTGCTTTGGCATCCATCTCTTGAAGAAGACTATCCAGAGAGATGAGA 914

Db 977 TGATCGATGCTCATGAGACCCAGATGTT---CGCGCATACCGGGGACCATGAGGA 1033

Qy 915 AGCAGCTGGCGGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTGGCATCTCAATCCG 974

Db 1034 AGTTGGTCGGTGACCGGCTGCCCAATTCACAGCCGAGCAGTGGAGTTGTGA---GG 1090

Qy 975 GAGAGCCGACCTTCTACGGCATTAATTACTACACATCCCATGTTGC 1020

Db 1091 GATCCTACGACTTCATCGGCTCCTCACTACTACACACCACTACGCG 1136

RESULT 13

AK120430 2148 bp mRNA linear PLN 29-OCT-2003

LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:J013095M12, full insert sequence.

DEFINITION

ACCESSION AK120430

VERSION AK120430.1 GI:37990053

KEYWORDS FLI_CDNA; CAP trapper.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eriarthroideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS

1 The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team; Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohta, Y., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Ootomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurokawa, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niki, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN; Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kawaga, I., Kondo, S., Konno, H., Miyazaki, A., Oota, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE

AUTHORS

TITLE Collection, mapping, and annotation of 28k full-length cDNA clones from japonica rice

JOURNAL unpublished

REFERENCE 3 (bases 1 to 2148)

AUTHORS Kikuchi, S.

TITLE Direct Submission

JOURNAL Submitted (31-JAN-2003) Shohei Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 32k full-length cDNA clones from japonica rice.

URL: [FEATURES

source

ORIGIN

Query Match 11.1%; Score 221.6; DB 8; Length 2148; Best Local Similarity 55.9%; Pred.No. 1.3e-34; Matches 529; Conservative 0; Mismatches 399; Indels 18; Gaps 5;

Qy 81 TGCCCAACGACTTTGATGGGGCTTGGCAACGCGCCCTACAGATGGAAGCGCCGTCA 140

Db 203 TCCCGAGAGACTTTCATCTTCGGACCGCGCTCCGCGCTTATCATGTAAGAGCGCTGTGA 262

Qy 141 AAGAAGGTGCGCGCGCCCTTCATCTTGGAGACGTAATGCACTGGAACCATCGCGCA 200

Db 263 ACGAAGCGCGCGAGAGCGCGAGCATCTGGAGACACGTACCTCCATTCAGGCGAAAGTTG 322

Qy 201 CCAAGCGCGCGCAACCGCGATGGCGTGGATACATACACACGCTCGATGAGAGACTTTC 260

Db 323 AAGATGTGACGACCGGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 382

Qy 261 ATCTCTTGACCAAGTACGCGCGCAAGAGCCCTACCGCTTCTCTTGTGCTGCGGATGA 320

Db 383 ACTTGTGACCGACATGACATGAGAGCGCTTCGCTTCTCCATGCTTGAAGAGATGC 442

Qy 321 TTCCCTCGCGCGCGAGCTGATCCGCTGAACGAGAGGAATTGATTTTACAGCAAC 380

Db 443 TGCAATGATGAACCATCACTGAGAGGATCAACAAAGAGGAGATGCTTCTTACACAGCC 502](http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project_Team_Kikuchi_S_Satoh_K_Nagata_T_Kawagashira_N_Doi_K_Kishimoto_N_Yazaki_J_Ishikawa_M_Yamada_H_Ooka_H_Hotta_I_Kojima_K_Namiki_T_Ohneda_E_Yahagi_W_Suzuki_K_Li_C_Ohta_Y_Shishiki_T_Yamamoto_M_and_Nakahama_Y_FALS_Genome_Sequencing_and_Analysis_Group_Ootomo_Y_Iida_Y_Fujimura_T_Ikeda_R_Ishibiki_J_Kawamata_M_Kobayashi_M_Kodama_T_Kurosaki_T_Kusumegi_T_Lu_M_Masuda_H_Miura_J_Mizuno_K_Narikawa_R_Nikura_J_Oka_M_Ryu_R_Sugano_S_Sugiyama_A_Suzuki_Y_Tsunoda_Y_Ueda_M_Xie_Q_Yokomizo_S_Yoshimura_A_Matsubara_K_and_Murakami_K_Genome_Exploration_Research_Group_in_Riken_Genomic_Sciences_Center_and_Genome_Science_Laboratory_in_Riken_Adachi_J_Aizawa_K_Akimura_T_Arakawa_T_Carninci_P_Fukuda_S_Hanagaki_T_Hara_A_Hashidume_W_Hayashida_K_Hayatsu_N_Hiramoto_K_Ishiohara_T_Hori_F_Iida_J_Imamura_K_Imotani_K_Ishii_Y_Itoh_M_Kagawa_I_Kanagawa_S_Kato_H_Kawai_J_Kishikawa_Hirozane_T_Kojima_Y_Kondo_S_Konno_H_Kouda_M_Koya_S_Kurihara_C_Matsuyama_T_Miyazaki_A_Murata_M_Nakamura_M_Nishi_K_Nomura_K_Nunaseki_R_Ohno_M_Oota_N_Ota_Y_Saitoh_H_Sakai_C_Sakai_K_Sakazume_N_Sano_H_Sasaki_D_Sato_K_Shibata_K_Shinagawa_A_Shiraki_T_Sogabe_Y_Tagami_M_Tagami-Takeda_Y_Tagawa_A_Takahashi_F_Takaku-Akahira_S_Takemi-Takeda_Y_Tanaka_T_Tomaru_A_Toya_T_Waki_K_Yasunishi_A_and_Hayashizaki_Y_Location/Qualifiers 1. 2148 /organism=)

```

QY 381 TGAATGACCCCTCTTGGAGCGGGGTATCAAGCCTTGGGTGATCTTGATCACTGGGAT 440
DB 503 TGAATTAACGAGGTCAATATCTAGAGGGTTGAAGCCATTGTGCACCATCTTCCATTTTGACA 562
QY 441 TGGCTAGAGGGCTTTCAGATCGCTATGAGAGGCTGAGCTCAACGTGAAGAGAGTCCAGCTGG 500
DB 563 CCCACACAGGCTCTGAGAGACAATAACGAGCTTCCCA---GTGAATAACATTGTGAAG 619
QY 501 ACTTTGACGGTATGCGAGGTTGTCTTTGAACCTTTTGGGAGCCAGTCCAGAACCTGGA 560
DB 620 ATTTCGTGAGCTATGCGAGCGTGTGCTTCGTGATTCGGCGAGCCGGGTAAAGTCAAGA 679
QY 561 TCCCATCAACGAGMCCCTGATTCAGGCACTATGATATGACACCGGAGCAACGCGCC 620
DB 680 ACGGTTCAACGAGCCCATATCTTCTGCGCCGCGGCTACGGCTCCGGACCAAGGCC 739
QY 621 CCGGAGAGAGC-----AGCATTAACAGCACTTCACCGAGGGCAACACTGCACTGAGC 674
DB 740 CCGGCGCTGCTGCGCGGTAGCTTCCAAAGAAATCGCCCGCGGCACTCCGGCAACGAGC 759
QY 675 CGTGGCTGCTGGAAGAGCCCAATCATAGAGCCATGCGCGCGCTGGCGCTTACAGCA 734
DB 800 CGTACGTGCTGCGGCAACCTGCTGCTGCGCCAGCGCGGCTGCTGCTTACCGCC 859
QY 735 GGGACTTTGCGCCCTCGCAAAAGGGCCAGATCGGATCTGCTCAACGCGGCACTACTATG 794
DB 860 AAAAGTACAGGCGACCGCAAGGGGAGATCGGCACTACCGAGTGTGCACTGCTTG 919
QY 795 AGCCCTTGGAGACGCAATGAGCTTGGGAGCAAGAGGCTGCTGAGCGACGATGGAATTC 854
DB 920 TGCCCT---ACAGGAGCGCGCGCCGCGCAAGGAGCGGCTGAGCGGAGCTTCACCTTCA 976
QY 855 ACATTTGCTGTTTGGCCATTCCTCTTGAAGAAGAGATATCCAGAGAGCATGAGA 914
DB 977 TGTAGGATGTTATGAGACCGGATCGTGT---CGGAGATACCGGCGCACTTGAAGA 1033
QY 915 AGCAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTCGCACTCTCAATGCG 974
DB 1034 AGTTGTGCTGAGACCGGCTGCGCAAAATTCACAGCGAGCACTCGAGATGTGTA---GG 1090
QY 975 GAGAGACCGACTTCTACGGCATGAAATCTACATCCCACTTGGC 1020
DB 1091 GATCTCACTGACTCATCGGCTCACTCACTACTACCACTACCACTACGC 1136

RESULT 14
AK066710 1811 bp mRNA linear PLN 24-JUL-2003
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:J013074A20, full
DEFINITION insert sequence.
ACCESSION AK066710
VERSION AK066710.1 GI:32976728
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Euphorbiaceae; Oryzaceae; Oryza.
REFERENCE
1 The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team;
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Kojima,K., Namiki,T., Ohneda,E., Yahagi,M., Suzuki,K., Li,C.,
Ohneda,E., Hori,F., Iida,J., Imamura,K., Imotani,K.,
Shisshiki,T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group; Otsu,Y., Murakami,K.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Nishikura,J.,
Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,S., Miura,J.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN;
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arikawa,T., Fukuda,S.,
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,

```

Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Otsu,N., Ota,Y.,
 Saito,R., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,
 Yoshino,M. and Hayashizaki,Y.
 Collection, mapping, and annotation of over 28,000 cDNA clones from
 japonica rice
 Science 301 (5631), 376-379 (2003)

JOURNAL MEDLINE
 PUBMED
 12869764
 2 (bases 1 to 1811)

REFERENCES
 Adachi,J., Aizawa,K., Akimura,T., Arikawa,T., Carninci,P., Doi,K.,
 Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashidume,W.,
 Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
 Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
 Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
 Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
 Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
 Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
 Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
 Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
 Namiki,T., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K.,
 Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
 Otsu,N., Ota,Y., Otsu,Y., Otsu,R., Satoh,K., Sakai,C., Sakai,K.,
 Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Sugano,S.,
 Shingawa,A., Shiraki,T., Shisshiki,T., Sogabe,Y., Sugano,S.,
 Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
 Tagawa,A., Takehashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
 Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
 Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
 Yoshimura,A.

TITLE
 JOURNAL
 Direct Submission
 Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
 tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT
 This clone is one of the 28k full-length cDNA clones from japonica
 rice.

URL: http://cdna01.dna.affrc.go.jp/cDNA/
 NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
 Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
 Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
 Ohneda,E., Yahagi,M., Suzuki,K., Li,C., Ohtsuki,K., Shisshiki,T. and
 Yamamoto,M.

FAIS Genome Sequencing & Analysis Group: Otsu,Y., Iida,Y.,
 Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
 Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
 Mizuno,K., Narikawa,R., Nishikura,J., Oka,M., Ryu,R., Sugano,S.,
 Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
 Yoshimura,A., Matsubara,K. and Murakami,K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
 Akimura,T., Arikawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
 Hara,A., Hashidume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
 Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
 Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Otsu,N.,
 Ota,Y., Satoh,K., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
 Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,
 Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
 Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
 Yasunishi,A. and Hayashizaki,Y.

FEATURES
 source
 1..1811
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultiVar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="J013074A20"

ORIGIN

Query Match 10.8%; Score 215.8; DB 8; Length 1811;
 Best Local Similarity 51.3%; Pred. No. 1.9e-33;
 Matches 714; Conservative 1; Mismatches 648; Indels 30; Gaps 8;

QY 81 TGCCCAACGACTTGAATGGGCTTCGCAACGGCCGCTTACAGATGAAAGGCGCTCA 140
 DB 135 TCCCAAGGACTTCATCTTCGGGACAGAGTTTCAGCATCTATCAGTAGAGGCGCTTACA 194
 QY 141 AAGAAGTGGCCGGGCGCGCTTCATCTGGGACAGTACGTGCACTGAGCACTCGGCA 200
 DB 195 AAGAAGGGGCAAGGCTTCTAGGCTCTGGGACAACTTCACTACATTCAGTAAATTT 254
 QY 201 CCAACGGCGCAACGGCGATGTGGCTTGCATCACTCAACCGCTACGATAGAGACTTGG 260
 DB 255 TAAACAATGATATACGGGATGTGGCAATGACTTCTATCAGCATATACAAGAGATGTGA 314
 QY 261 ATCTTTGACCAAGTACGGGCGCAAGGCTTACGGCTTCTCTGTGTGTGTGGATCA 320
 DB 315 GCCTCTGGAAGACATGAATGATGCTTCCGGTCTTCATTTGCGTGAACAGATCC 374
 QY 321 TTCCCTCGGCGGAGGCTGATCCCGTCAACGAGAGGGAATGATTTTACAGCAAC 380
 DB 375 TGCCAAATGATCTTATAGTGAAGAAATTAACAAAGAGGCTTCTTCTACCAACGCT 434
 QY 381 TGAATGACGCTTGTGAGCGGGGTATCAAGCTTGGGTGACTTTGTACACTGGATC 440
 DB 435 TGATCATGATGTATATGCAAAAGGATGATCCATTTGTCACTATCTTCCATCGGACA 494
 QY 441 TGCTCAGGCGCTTACAGATCGCTATGAGCGTGGCTCAACGTGAGAGGTCCAGCTGG 500
 DB 495 CCCCCTTGGCTTGAAGAAAGCAATACGAGGATTTCTCA--GTGAAGACATAGTGAAG 551
 QY 501 ACTTTGAGCGGTATGCGAGGTGTGCTTTGACCTTTTGGGAGACCGAGTCAAGACTGGA 560
 DB 552 AATACTGGAATTCGCGAGGAGTGTCTTTCGCAATTCGCGAGCGGTGTCAAGATCTGGA 611
 QY 561 TCACCATCAACGAGWCCCTGATTCAGGCACTATGATATAGCAGCGGAGCAACGCC 620
 DB 612 CCACATTTAATGAGCCATTCAATATAGCCCTACGCTACGCAAGGCGCTGTTCGCC 671
 QY 621 CGGGCAGAGACGCA-----TTAACAAGCACTTCACGAGGCGACACTGCGCATGAGC 674
 DB 672 CGGGACGATGCTCTTCAATGATTTCCAAATCATGCGGCGTGGGATCCAGTGCAGC 731
 QY 675 CGTGGCTCGCTGAAAGGCCCAATCATGAGCCATGCCCCGCGTGGCCGTCTACGCA 734
 DB 732 CTAACCTGTGGGACACCAATCATCACTCTCCACGCTGCGCTGTCAAGCTTTACGCA 791
 QY 735 GGGACTTTGCGCCCTCGCAAAAGGGCCAGATCGGATCTGCTCAACGGGCACTACTATG 794
 DB 792 CCAAGTACAGCCCAACAGAGGAGAGATCGGATGTGTGTGTATACCACTGTGTTCC 851
 QY 795 AGCCCTGGGACAGCAATGAGCTTCGGGACAAAGAGCTGTGAGCGACGATGGATTTTC 854
 DB 852 TGCGGTACGACCACTCCGAGCT--GACCGTGGCGCTGTGCAACGAGGCTTACACTTCA 908
 QY 855 ACATGTGCTGTGGTTCGCAATCCATCTTCTTGAAGAGACTATTCAGAGCATGGAAGA 914
 DB 909 TCTATGGGTGTTCATGAGACCTTA--TGTGTGCAATGTGTACTACCGAGCACTGTGAG 965
 QY 915 AGCAGCTGGGCGAGAGGCTTCAGACCTCACTCCGGGACTTTGGCATCTCTCAATCCG 974
 DB 966 GTTGGCTGGCAATCGGCTGCGAGATTCACGCTTGAACAGTGGCCATGTGTAGAGGCT 1025
 QY 975 GAGAGACGACTTTACGGGATGAAATTAACAATCCAGTTTGCAGCGCACTAGACG 1034
 DB 1026 CTTA---GCACTTCATCGAGTTAATTAACAACCACTACTAGGCTAAGAGTATACCGC 1082
 QY 1035 GTCCCGTCCCGAGACG-----ACTATCTCGGCGCATTCATGAGACCAAGAGATA 1088
 DB 1083 CGCTTAATCTCAACGAGCTATCTTAAGACCTTCGCAACCGGCGCAACACCGGCTTCC 1142
 QY 1089 AGGACGGACCGCGCTTGGGAGAGAGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCA 1148

DB 1143 GTAATGGCAAAACCATCGGTCCACAGAAATTTACACCGATCTTCTTCACTACCTCCAG 1202
 QY 1149 TGTTCCGGAAGACTTCCTCCCGGGGTATACGGCTGTACGCGCAAGCCCATATATACCG 1208
 DB 1203 GTCTCGTGAGCTTCCTCTTACCAAGAGAGATTAACAACCCGA---CATCTATG 1259
 QY 1209 AGAAGGATGCCCGTCCCTGAGAGAGAACATGACGTGCGAGAGGCGGTCAAGACC 1268
 DB 1260 TTACAAAAAGGCAATCATGAGGTATACAGACACACTGCGCAAGGCGCTCAAGATG 1319
 QY 1269 CTTCCGCAATCGGTACTTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1328
 DB 1320 GACACAGAGATCGAGTTCCACTCAAAAGCACTGAGTTGCTCAACAGCCAT---CAAGA 1376
 QY 1329 ACGGCTGCTGCTCAAGGGGTACTTTCGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1388
 DB 1377 ATGGGCTGAACGTGAAGGGCTTACTTCAAGCATTTGATGACTGCTTGTGAGTGGGCTG 1436
 QY 1389 ATGCTACGGAACCGATTCGGGCTGACGTTCAAGACTTACACACCTCAAGCGACGC 1448
 DB 1437 ACGGCTACCTTGGACAGGTTGCGCTCATTTACGTCAACGCAAGACGCTTACGCTACC 1496
 QY 1449 CCAAGAAGTCTGC 1461
 DB 1497 GCAAGAGTCCAGC 1509

RESULT 15
 AX652977 1503 bp DNA linear PAT 22-MAR-2003
 LOCUS AX652977
 DEFINITION Sequence 2847 from Patent WO03000898.
 ACCESSION AX652977
 VERSION AX652977.1 GI:29155791
 KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriatloideae; Oryzaceae; Oryza.
 REFERENCE
 1 Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
 Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
 TITLE Plant genes involved in defense against pathogens
 JOURNAL Patent: WO 03000898-A 2847 03-JAN-2003;
 FEATURES
 Location/Qualifiers
 1..1503
 /organism="Oryza sativa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:4530"

ORIGIN

Query Match 10.8%; Score 214.6; DB 6; Length 1503;
 Best Local Similarity 51.0%; Pred. No. 3.4e-33;
 Matches 720; Conservative 1; Mismatches 660; Indels 30; Gaps 8;

QY 81 TGCCCAACGACTTGAATGGGCTTCGCAACGGCCGCTTACAGATGAAAGGCGCTCA 140
 DB 95 TCCCAAGGACTTCATCTTCGACAGAGTTTCACACACTTATCAGTATGAGGGGCGCTTACA 154
 QY 141 AAGAAGTGGCCGGGCGCGCTTCATCTGGGACAGTACGCTGCACTGAGCACTCGGCA 200
 DB 155 AAGAAGGGGCAAGGCTTCTAGGCTCTGGGACAACTTCACTCACTTCCAGTAAATTT 214
 QY 201 CCAACGGCGCAACGGCGATGTGGCTTGCATCACTACACCGCTACGATGAGAGACTTGG 260
 DB 215 TAAACAATGATATACGGGATGTGGCAATGACTTCTATACCGATACAAGAGAGATGTGA 274
 QY 261 ATCTTTGACCAAGTACGGGCGCAAGGCTTACGCTTCTGCTGTGTGTGTGGATCA 320
 DB 275 GCCTCTGGAAGACATGAATGATGCTTTCGGTCTTCTCAATGCTGTGACCAAGATCC 334

```
OY 321 TTCCCTCGGCGGCGAGCTGGATCCCTCAACGAGAGGAATTGATTTTACAGCAAC 380
Db 335 TGCCTAAATGATTCCTTATGAGAGAAATAACAAAGAGGGGTGCTTTCTACAAACGCT 394
OY 381 TGATTGACGCCCTGTGAGCGGGGTATCAAGCCTTGGGTGACTTTGTACACTGGGATC 440
Db 395 TGATCAATGATGTATGACAAAGAGGATGATCCATTGTTCACATCTTCCACTGGGACA 454
OY 441 TGCCCTCAGGCGCTTACGATCGTATGAGAGCTGGCTCAACGTGAAAGAGGTCCAGCTG 500
Db 455 CCCCCTTGCTCTGAAAGCAAAATACGAGGATTCCTCA--GTGAAGACATAGTGAAG 511
OY 501 ACTTGACCGGTATGCCAGGTGTGCTTTGAAACGTTTGGGGACCGAGTCCAGAACTGGA 560
Db 512 AATAGTGAGCTTGGGAGGTGTGCTTCCGAGTTCGGACCGTGTCAAGTACTGGA 571
OY 561 TCACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACGGGAGCAACGCC 620
Db 572 CCACTTTAATGAGCATTTCAATATAGCGCTACGGCTACGGCAAGGGCGTGTGCGC 631
OY 621 CGGCGAGAGCAACA-----TTAACAAGCACTCCACGAGGCAACCTGCCACTGAGC 674
Db 632 CGGGAAGATGCTCTTCAATATGTTTCCAGTATGCGGCTGCTGACTTCAGTCGCAAGC 691
OY 675 CGTGCGCTGCGGAAAGGCCAGATCATGAGCAATGCCGCGCGTGGCCGTCTACAGCA 734
Db 692 CTTACTTGAGGACACACATCCACTCTCCACAGCTGCGCTGTCCAGCTTTACGCA 751
OY 735 GGGACTTTCGCCCTCGCAAAAGGGCCAGTCGCGATCTCGTCAACGGCGACTACTATG 794
Db 752 CCAAGTACAGCCACACAGAAAGGAGCAGATCGCATGTGTGTCAACCACTGTGTCG 811
OY 795 AGCCCTGGAGCAAGCATGAGCCTCGGACCAAGAGGCTGTGAGCGACGAGATGAAATTTC 854
Db 812 TGCCCTGACCAACTCCGACGCT--GACCGTGCGCTGTGCAACGAGCCTAGACTTCA 868
OY 855 ACATTGCTGTGTTGCCAATCCCATCTTCTTGAAAGAGACTATCCAGAGCATGAAGA 914
Db 869 TCTATGGGTGTTCATGAGCCCTA---TGATCATGTGACTACCAAGCACATGAGAG 925
OY 915 AGCAGCTGGGAGAGAGCTTCCAGCCCTGATCCCGGAGCTTTCGCAATGCCG 974
Db 926 GTTGGCTCGGCAATCGGCTCGCAGAGTTCAAGCTGAACAGTCGCGATGTGAAGGCT 985
OY 975 GAGAGCCGACTTTCACGCGATGAATTACTACACATCCCAATTGCGCGCACCTAGACG 1034
Db 986 CCTA--CGACTTTCATCGAGTTAATTACTACACACACTACTACGTAAAGATATACGCG 1042
OY 1035 GTCCCGTCCCGA-----GACGGAATCTCGGCGCCATCCATGAGCACAGAGAAATA 1088
Db 1043 CGCTTAATCTCAACGAGTATCTTACGACTCGACCAACCGCGCAACACACCGGCTTCC 1102
OY 1089 AGGACGGCAGCCCGCTTGGCGAGAGAGGGGCTCGGCTCGGCTCGGCCGAGACA 1148
Db 1103 GTAATGCAAAACCATCGGTTCACAGAAATTACACGATCTTCTTCACTAACCTCCAG 1162
OY 1149 TGTTCGGAAGCATCTGCCCCGGGTGTAGGCTGTACGCGCAAGCCCATCTACATCACCG 1208
Db 1163 GTCTCGTGAGCTCTCTCTTACACCAAGAGAGATTAACAACCCGA---CCATCTATG 1219
OY 1209 AGAACGATGCGCGTCCCTGAGAGAGAAATGACGTGCGAGAGGCGCTCAAGACC 1268
Db 1220 TTACAGAAAACGCGATGATGAGGTACAAACGACACATGCGCAGAGGCGCTCAAGATG 1279
OY 1269 CCTTCGCAATCCGSTACTTGAATCGCATTTGGAATGGAATTCAGAGCCATTACCAGG 1328
Db 1280 GACACAGGATCGAGTTCATCAAAAGCACTGCGATTCGTAACACAGCCAT---CAGA 1336
OY 1329 ACCGCGTGTGTCGAAGGGGTACTTTGCGGGCGTGTGCTCGATTAATTGAAATGCTCAG 1388
Db 1337 ATGGGGTGAAGTGAAGGGCTACTTCACTGCGCATTCATGAGACTGCTTGAAGGGGTG 1396
OY 1389 ATGGCTAGGAGCCAGATTGGCGGTCACTTCAAGACTACACCACTCACCCTCAAGCGCAGC 1448
```

```
Db 1397 ACGGCTACTTGACAGAGTTGCGGCTCATCTACGTGACCGCAAGCGCTCAAGCGCTACC 1456
OY 1449 CCAAGAACTTCGCCCTGCTCTCAAGACAT 1479
Db 1457 GCAAGGATCCAGCTACTGGATCGAAGACTT 1487
```

Search completed: March 28, 2004, 20:51:58
Job time : 7890.13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2004, 19:13:22 ; Search time 788.658 Seconds
(without alignments)
10724.755 Million cell updates/sec

Title: US-10-026-140-1
Perfect score: 1991
Sequence: 1 agccagtcgcacacagcag.....tttcaaaaaaaaaaaaaa 1991

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseq19908:*
2: geneseq19908:*
3: geneseq20009:*
4: geneseq2001a:*
5: geneseq2001b:*
6: geneseq2002a:*
7: geneseq2003a:*
8: geneseq2003b:*
9: geneseq2003c:*
10: geneseq2004a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	ID	Description
1	1976.6	99.3	1990	9	ADD22902 Full leng
2	1454.2	73.0	1455	9	ADD22904 T. reesei
3	598	30.0	735	3	AAFI4935 Trichoder
4	266.6	13.4	1933	3	AAFI4882 Trichoder
5	246	12.4	1404	7	ADA70881 Rice gene
6	214.6	10.8	1503	7	ADA69524 Rice gene
7	205.4	10.3	575	7	ABZ53130 Aspergill
8	182.8	9.2	1734	6	ABZ13477 Arabidops
9	182.8	9.2	1734	7	ADA67833 Arabidops
10	178.6	9.0	1533	7	ADA70381 Rice gene
11	178.2	9.0	2016	4	AA521370 Human CDN
12	178.2	9.0	2016	6	ABL88245 Human PRO
13	178.2	9.0	2016	7	ABL95734 Human ang
14	178.2	9.0	2016	7	ACD23979 Human hum
15	178.2	9.0	2016	7	ACA67120 CDNA enco
16	178.2	9.0	2016	7	ACA03729 CDNA enco
17	178.2	9.0	2016	7	ABX89267 DNA enco
18	178.2	9.0	2016	7	ACD41921 Human sec
19	178.2	9.0	2016	8	ACA04150 Human CDN
20	178.2	9.0	2016	8	ADA45772 Human hum
21	178.2	9.0	2016	8	ADA76203 Human PRO
22	178.2	9.0	2016	8	ADA18853 Human PRO
23	178.2	9.0	2016	8	ADA61476 Homo sapi

24	178.2	9.0	2016	8	ADB19261	Novel hum
25	178.2	9.0	2016	8	ADB27802	CDNA enco
26	178.2	9.0	2016	8	ADA6281	Novel hum
27	178.2	9.0	2016	8	ADB15845	Human PRO
28	178.2	9.0	2016	8	ADA7631	Human PRO
29	178.2	9.0	2016	8	ADA67426	Human PRO
30	178.2	9.0	2016	8	ADB30433	CDNA enco
31	178.2	9.0	2016	8	ADA85729	Novel hum
32	178.2	9.0	2016	8	ADA69941	Human PRO
33	178.2	9.0	2016	8	ADA79245	Human PRO
34	178.2	9.0	2016	8	ADA87384	Novel hum
35	178.2	9.0	2016	8	ADB16586	Human PRO
36	178.2	9.0	2016	8	ADA91678	Novel hum
37	178.2	9.0	2016	8	ADB14741	Human PRO
38	178.2	9.0	2016	8	ADB18702	Novel hum
39	178.2	9.0	2016	8	ADA93917	Human PRO
40	178.2	9.0	2016	8	ADB19813	Novel hum
41	178.2	9.0	2016	8	ADB13125	Human PRO
42	178.2	9.0	2016	8	ACD98550	Novel hum
43	178.2	9.0	2016	8	ADA74379	Human PRO
44	178.2	9.0	2016	8	ADB24612	Human PRO
45	178.2	9.0	2016	8	ADA82136	Human PRO

ALIGNMENTS

RESULT 1	ADD22902	ADD22902 standard; CDNA; 1990 BP.
XX	XX	
AC	ADD22902;	
XX	XX	
DT	15-JAN-2004	(first entry)
XX	XX	
DE	Full length T. reesei bg15 beta-glucosidase CDNA.	
XX	XX	
KW	BGL5; beta-glucosidase; fungus; ethanol; sugar;	
KM	fermentative microorganism; endoglucanase; cellobiohydrolase; wine making;	
KW	aroma; detergent; softening agent; cotton; wood pulp; gene; ss.	
XX	XX	
OS	Hypocrea jecorina.	
XX	XX	
PH	Key	Location/Qualifiers
FT	CD5	62..1516
FT		/*tag= a
FT		/product= "BGL5 beta galactosidase"
FT		/transl_except= (pos:98..100,aa:Gln)
FT		/transl_except= (pos:551..553,aa:Gln)
FT		/transl_except= (pos:572..574,aa:Xaa)
FT		/note= This codon has an apparent 1 nucleotide deletion which alters the reading frame. Xaa= unknown.
XX	XX	
US	US2003114330-A1.	
PD	19-JUN-2003.	
XX	XX	
PF	18-DEC-2001; 2001US-00026140.	
XX	XX	
PR	18-DEC-2001; 2001US-00026140.	
XX	XX	
PA	(DUNN/) DUNN-COLEMAN N.	
PA	(GOED/) GOEDEGEBOUR F.	
PA	(WARD/) WARD M.	
PA	(YAO/) YAO J.	
XX	XX	
PI	Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;	
XX	XX	
DR	WPI; 2003-810909/76.	
DR	P-PSDB; ADD22903.	
XX	XX	
PT	New substantially purified BGL5 polypeptide with the biological activity	

PT of beta-glucosidase, useful for degrading biomasses to ethanol.

XX Example 1; Fig 1; 21pp; English.

CC The invention discloses a substantially purified BG45 polypeptide with
CC the biological activity of a beta-glucosidase. Also claimed is an
CC isolated polynucleotide derived from a fungal source, which comprises a
CC nucleotide sequence encoding an enzyme having beta-glucosidase activity,
CC a *Trichoderma reesei* bg45 nucleic acid coding sequence, or its
CC complement, an expression construct including the *T.reesei* bg45 cDNA
CC sequence, a purified recombinant enzyme having beta-glucosidase activity,
CC a recombinant host cell comprising a deletion or insertion or other
CC alteration in the bg45 gene which inactivates the gene and prevents BG45
CC polypeptide production, an antisense oligonucleotide, a detergent
CC composition comprising the polypeptide and expressing a heterologous
CC polypeptide having beta-glucosidase activity in an *Aspergillus* sp. BG45
CC is useful for producing ethanol, which involves contacting a biomass
CC composition with an enzymatic composition comprising the polypeptide to
CC yield a sugar solution, adding to the sugar solution a fermentative
CC microorganism and culturing the fermentative microorganism under
CC conditions sufficient to produce ethanol, where the biomass composition
CC may be optionally pretreated. The method further involves the addition of
CC at least one endoglucanase or cellobiohydrolase. The pretreatment is with
CC a dilute acid. BG45 is useful in wine making for enhancing the potential
CC aroma of the finished wine product. The detergent is useful as softening
CC agent and also useful for improving the feel of cotton fabrics and for
CC degrading wood pulp into sugars. The sequence presented is the full
CC length *T. reesei* bg45 cDNA.

XX Sequence 1990 BP; 439 A; 611 C; 504 G; 434 T; 0 U; 2 Other;

Query Match 99.3%; Score 1976.6; DB 9; Length 1990;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1989; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
QY 1 AGCCAGGTGGAACGAGAGGAGCGAGCTACAGAGAAATCAACCAATAGCTCAA 60
DB 1 AGCCAGGTGGAACGAGAGGAGCGAGCTACAGAGAAATCAACCAATAGCTCAA 60
QY 61 AATGCCAGTGTGCTAGCTGTGCCAAGACTTTGAATGGGGCTTGGCAAGGGCGGCTTA 120
DB 61 AATGCCAGTGTGCTAGCTGTGCCAAGACTTTGAATGGGGCTTGGCAAGGGCGGCTTA 120
QY 121 CCAAGTTCGAAGCGCCGTCGAAGAAAGTGGCCGCGCCGCTCACTTGGGACAGTACTG 180
DB 121 CCAAGTTCGAAGCGCCGTCGAAGAAAGTGGCCGCGCCGCTCACTTGGGACAGTACTG 180
QY 181 CCACCTGGAGCATGCGCAACCAACGCGGCAAGGGGATGTGGCTTGCATCACTACCA 240
DB 181 CCACCTGGAGCATGCGCAACCAACGCGGCAAGGGGATGTGGCTTGCATCACTACCA 240
QY 241 CCGCTACGATGAGGACTTTGATCTCTTGAACCAAGTACGCGCAAGGCTTACGCTTCTC 300
DB 241 CCGCTACGATGAGGACTTTGATCTCTTGAACCAAGTACGCGCAAGGCTTACGCTTCTC 300
QY 301 CTGTGTGCTGTCGGGATCATTCCTCTCGGCGGCGAGGCTGGATCCCTCAACGAGAGGG 360
DB 301 CTGTGTGCTGTCGGGATCATTCCTCTCGGCGGCGAGGCTGGATCCCTCAACGAGAGGG 360
QY 361 AATTGAGTTTACAGCAAACTGATTGAGCCCTGTGAGGGGGGTATCAACGCTTGGGT 420
DB 361 AATTGAGTTTACAGCAAACTGATTGAGCCCTGTGAGGGGGGTATCAACGCTTGGGT 420
QY 421 GACTTTGTACCACTGGGATCTGCTCAAGCGCTTCAAGATCGCTATGAGAGGCTGCTCAA 480
DB 421 GACTTTGTACCACTGGGATCTGCTCAAGCGCTTCAAGATCGCTATGAGAGGCTGCTCAA 480
QY 481 CGTGAAGAGGTCCAGCTGGAATTTGAGCGGTATGCAAGGTGTGCTTTGAACGTTTGG 540
DB 481 CGTGAAGAGGTCCAGCTGGAATTTGAGCGGTATGCAAGGTGTGCTTTGAACGTTTGG 540
QY 541 GGAACGAGTCCAGAACTGATCAACATCAAGAACCTGTGATTGAGGCAATCTATGATA 600
DB 541 GGAACGAGTCCAGAACTGATCAACATCAAGAACCTGTGATTGAGGCAATCTATGATA 600
```

```
DB 541 GGACCGAGTCCAGAACTGATCAACATCAAGAACCTGTGATTGAGGCAATCTATGATA 600
QY 601 TGCCACCGGAGCAACCGCCCGGAGAGAGAGATTTAAACAGACTTCACCGAGGCAA 660
DB 601 TGCCACCGGAGCAACCGCCCGGAGAGAGAGATTTAAACAGACTTCACCGAGGCAA 660
QY 661 CACTGCACTGAGCGCGTGTGCTGTGAAAGGCCAGATATGAGCCATGCCCCCGCT 720
DB 661 CACTGCACTGAGCGCGTGTGCTGTGAAAGGCCAGATATGAGCCATGCCCCCGCT 720
QY 721 GGCGGTCTAAGAGAGGAGCTTTGCGCCCTCGCAAAAGGGCCAGATGGGACTTCTCAA 780
DB 721 GGCGGTCTAAGAGAGGAGCTTTGCGCCCTCGCAAAAGGGCCAGATGGGACTTCTCAA 780
QY 781 CGGCGACTACTATGAGCCTTGGGACAGCAATGAGCCTTGGGACAGAGGCTTGAAGCG 840
DB 781 CGGCGACTACTATGAGCCTTGGGACAGCAATGAGCCTTGGGACAGAGGCTTGAAGCG 840
QY 841 ACGGATGAATTTCAATTTGGCTGTGTTTGCAATCCATCTTCTTGAAGAGACTATCC 900
DB 841 ACGGATGAATTTCAATTTGGCTGTGTTTGCAATCCATCTTCTTGAAGAGACTATCC 900
QY 901 AGAGAGCATGAAAGAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTC 960
DB 901 AGAGAGCATGAAAGAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTC 960
QY 961 CATCTCTAATGCGGAGAGAGAGCACTTTCTACGGCATGAATTTACTACATCCCAATTCC 1020
DB 961 CATCTCTAATGCGGAGAGAGAGCACTTTCTACGGCATGAATTTACTACATCCCAATTCC 1020
QY 1021 GGCGCACTTGAAGAGCTCCGCTCCCGAGAGCGGACTTCTCGGCGCATCAAGAGACCA 1080
DB 1021 GGCGCACTTGAAGAGCTCCGCTCCCGAGAGCGGACTTCTCGGCGCATCAAGAGACCA 1080
QY 1081 GGAAGATTAAGAGAGCGAGCCCGTTGGCGAGAGAGAGAGGCGCTCGCTGGCTCTG 1140
DB 1081 GGAAGATTAAGAGAGCGAGCCCGTTGGCGAGAGAGAGAGGCGCTCGCTGGCTCTG 1140
QY 1141 CCGGACATGTTCCGGAACATCTGCGCCGGGTGTACGGCTGTATAGGCAAGCCATCTA 1200
DB 1141 CCGGACATGTTCCGGAACATCTGCGCCGGGTGTACGGCTGTATAGGCAAGCCATCTA 1200
QY 1201 CATCACCGAAGAGATGCGGCTGCGCTGAGAGAGAGAACTATGACGTGAGAGAGCGGT 1260
DB 1201 CATCACCGAAGAGATGCGGCTGCGCTGAGAGAGAGAACTATGACGTGAGAGAGCGGT 1260
QY 1261 CAAGCAAGCCCTTCCGATCCGCTACTTTGACTTGGACTTGGACTGTGATTTCCAGGCCAT 1320
DB 1261 CAAGCAAGCCCTTCCGATCCGCTACTTTGACTTGGACTTGGACTGTGATTTCCAGGCCAT 1320
QY 1321 TACCCAGAGCGGCGTGTGTCTCAAGGGGATCTTTCGTGGGGGTGCTCGAATACTTGA 1380
DB 1321 TACCCAGAGCGGCGTGTGTCTCAAGGGGATCTTTCGTGGGGGTGCTCGAATACTTGA 1380
QY 1381 ATGTGAGATGAGCTACGAGACCCAGATTCGCGCTCAAGCTTCAAGAGCTACACCACTCTCAA 1440
DB 1381 ATGTGAGATGAGCTACGAGACCCAGATTCGCGCTCAAGCTTCAAGAGCTACACCACTCTCAA 1440
QY 1441 GGGCAGCGCCCAAGAGTGTGCTGCTCTCAAGAGCATGTTTGGCGCCCGGAGAGGCT 1500
DB 1441 GGGCAGCGCCCAAGAGTGTGCTGCTCTCAAGAGCATGTTTGGCGCCCGGAGAGGCT 1500
QY 1501 TAAAGTGGCGGATTAAGAAAGGAAATTTCTCTTGAATTCAGCTCTATGACATCTTCC 1560
DB 1501 TAAAGTGGCGGATTAAGAAAGGAAATTTCTCTTGAATTCAGCTCTATGACATCTTCC 1560
QY 1561 TCTCTTTTTCCTCCCTCCCTTGTCCCTCTCTCTACCTCATATTTCCCTCTATA 1620
DB 1561 TCTCTTTTTCCTCCCTCCCTTGTCCCTCTCTCTCTACCTCATATTTCCCTCTATA 1620
QY 1621 CCCCCCGCTTCTCTCAAGACCCCATGCTCTTGGCCCTTGGCCCTCTCTGTGCAATCT 1680
DB 1621 CCCCCCGCTTCTCTCAAGACCCCATGCTCTTGGCCCTTGGCCCTCTCTGTGCAATCT 1680
```

Oy	1681	GCCTTTATACAGCTTATAGGCTGTTTACTATGCTTTTCTTTTCTCTTCGT	1740
Db	1680	GCCCTTTATACAGCTTATAGGCTGTTTACTATGCTTTTCTTTTCTCTTCGT	1739
Oy	1741	CTGTCTGTCTCCCTGTCTATGTATACCTATCTGGCCCTTCGCTCATGGCAACAGATCT	1800
Db	1740	CTGTCTGTCTCCCTGTCTATGTATACCTATCTGGCCCTTCGCTCATGGCAACAGATCT	1799
Oy	1801	AGCAACAATTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCCATCAACGGCATC	1860
Db	1800	AGCAACAATTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCCATCAACGGCATC	1859
Oy	1861	AAAGCCCCATGTTTAGCCTCATGTTACATTTGCTATGTTATCTACATAGCCATTCACTA	1920
Db	1860	AAAGCCCCATGTTTAGCCTCATGTTACATTTGCTATGTTATCTACATAGCCATTCACTA	1919
Oy	1921	CCAGGCGAAGAGGCGACAGAGATCTCATTCGCTTAACTGTATATACGCTTTTTPAAAA	1980
Db	1920	CCAGGCGAAGAGGCGACAGAGATCTCATTCGCTTAACTGTATATACGCTTTTTPAAAA	1979
Oy	1981	AAAAAAAAAAAA 1991	
Db	1980	AAAAAAAAAAAA 1990	

XX	ADD22904	standard; cDNA; 1455 BP.
XX	ADD22904;	
XX	15-JAN-2004	(first entry)
XX	T. reesei bgl5 beta-glucosidase cDNA.	
XX	BGL5; beta-glucosidase; fungus; ethanol; sugar;	
XX	fermentative microorganism; endoglucanase; cellobiohydrolase; wine making	
XX	aroma; detergent; softening agent; cotton; wood pulp; gene; ss.	
OS	Hypocrea jecorina.	
XX	Key	Location/Qualifiers
XX	CDS	1..1455
FT	/*tag= a	
FT	/product= "BGL5 beta galactosidase"	
FT	/transl_except= (pos:37..39,aa:Gln)	
FT	/transl_except= (pos:511..513,aa:Xaa)	
FT	/transl_except= (pos:1219..1221,aa:Pro)	
FT	/note= "Xaa= unknown"	
XX	US2003114330-A1.	
XX	19-JUN-2003.	
XX	18-DEC-2001; 2001US-00026140.	
XX	18-DEC-2001; 2001US-00026140.	
XX	(DUNN/) DUNN-COLEMAN N.	
XX	(GOED/) GOEDEGEBOUR F.	
XX	(WARD/) WARD M.	
XX	(YAO/) YAO J.	
XX	Dunn-Coleman N, Goedegebuur F, Ward M, Yao J,	
XX	WPI; 2003-810909/76.	
XX	P-PSDB; ADD22903.	
XX	New substantially purified BGL5 polypeptide with the biological activity	
XX	of beta-glucosidase, useful for degrading biomass to ethanol.	
XX	Claim 2; Page 16-17; 21pp; English.	

XX The invention discloses a substantially purified BGL5 polypeptide with
CC the biological activity of a beta-glucosidase. Also claimed is an
CC isolated polynucleotide derived from a fungal source, which comprises a
CC nucleotide sequence encoding an enzyme having beta-glucosidase activity,
CC a Trichoderma reesei Bgl5 nucleic acid coding sequence, or its
CC complement, an expression construct including the T. reesei bgl5 cDNA
CC sequence, a purified recombinant enzyme having beta-glucosidase activity,
CC a recombinant host cell comprising a deletion or insertion or other
CC alteration in the bgl5 gene which inactivates the gene and prevents BGL5
CC polypeptide production, an antisense oligonucleotide, a detergent
CC composition comprising the polypeptide and expressing a heterologous
CC polypeptide having beta-glucosidase activity in an Aspergillus sp. BGL5
CC is useful for producing ethanol, which involves contacting a biomass
CC composition with an enzymatic composition comprising the polypeptide to
CC yield a sugar solution, adding to the sugar solution a fermentative
CC microorganism and culturing the fermentative microorganism under
CC conditions sufficient to produce ethanol, where the biomass composition of
CC may be optionally pretreated. The method further involves the addition of
CC at least one endoglucanase or cellobiohydrolase. The pretreatment is with
CC a dilute acid. BGL5 is useful in wine making for enhancing the potential
CC aroma of the finished wine product. The detergent is useful as softening
CC agent and also useful for improving the feel of cotton fabrics and for
CC degrading wood pulp into sugars. The sequence presented is the T. reesei
CC bgl5 cDNA.
XX
XX
SQ Sequence 1455 BP; 312 A; 444 C; 423 G; 274 T; 0 U; 2 Other;

Query Match	73.0%	Score 1454.2	DB 9	Length 1455
Best Local Similarity	100.0%	Pred. NO. 0		
Matches 1455	Conservative 0	Mismatches 0	Indels 0	Gaps 0

50	Sequence	1455 BP; 312 A; 444 C; 423 G; 274 T; 0 U; 2 Other;	
	Query Match	73.0%; Score 1454.2; DB 9; Length 1455;	
	Best Local Similarity	100.0%; Pred. No. 0;	
	Matches 1455; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY		62 ATGCCCGAGTGTCTAGCTCTGSCCAAGACTTTGAATGGGGCTTCGCAACGGCCGCTTAC	121
DB		1 ATGCCCGAGTGTCTAGCTCTGSCCAAGACTTTGAATGGGGCTTCGCAACGGCCGCTTAC	60
QY		122 CAGATCGAAGGCGCGCTCAAGAAAGAGTGGCGCGCGCCGCTCATCTGGGAGACGTACTGC	181
DB		61 CAGATCGAAGGCGCGCTCAAGAAAGAGTGGCGCGCCGCTCATCTGGGAGACGTACTGC	120
QY		182 CACCTGAGCCATCGCGCACCAACGAGCGCCACCGCATGTGGCTTGCATCACTAAC	241
DB		121 CACCTGAGCCATGTGGCCACCAAGGGGCCAACCGCATGTGGCTTGCATCACTAAC	180
QY		242 CGCTACGATGAGGACTTTGATCTTCTTGACCAAGTACGAGCGCCAAAGGCTTACCGCTTCTC	301
DB		181 CGTACGATGAGGACTTTGATCTTCTTGACCAAGTACGAGCGCCAAAGGCTTACCGCTTCTC	240
QY		302 TTGTGCTGCTGCGCGGATCATTCCTCCCTGGCGGCGAGCTGTGATCCCGTCAACGAGAGAGGA	361
DB		241 TTGTGCTGCTGCGCGATCATTCCTCCCTGGCGGCGAGCTGTGATCCCGTCAACGAGAGAGGA	300
QY		362 ATTGAGTTTTCACAGCAAACTGATTTGACGCGCTGTGTGAGGCGAGGATATCAAGCTTTGGGTG	421
DB		301 ATTGAGTTTTCACAGCAAACTGATTTGAGCGCCTGTGTGAGGCGAGGATATCAAGCTTTGGGTG	360
QY		422 ACTTTGTATCAACTGGGATCTTGCTCAGGCGCTTTCACGATGTGCTATGAGAGCTGAGCTCAAC	481
DB		361 ACTTTGTATCAACTGGGATCTTGCTCAGGCGCTTTCACGATGTGCTATGAGAGGCTGAGCTCAAC	420
QY		482 GTGGAAGAAGTCTCAGCTGTGACTTTTGAAGCGGATATCGAGAGTGTGTCTTTGAACGTTTGGG	541
DB		421 GTGGAAGAAGTCTCAGCTGTGACTTTTGAAGCGGATATCGAGAGTGTGTCTTTGAACGTTTGGG	480
QY		542 GACCGAGATCCGAAGCTGATCAGCATGAACGAMCCCTGGATATTACGGGCATCTATGATAT	601
DB		481 GACCGAGATCCGAAGCTGATCAGCATGAACGAMCCCTGGATATTACGGGCATCTATGATAT	540
QY		602 GCACAACGGGCAACAACGCCCGCGGAGAGAGCACTTAAACAAGCACTCCACGAGAGGCAAC	661
DB		541 GCACAACGGGCAACAACGCCCGCGGAGAGAGCACTTAAACAAGCACTCCACGAGAGGCAAC	600
QY		662 ACTGCGACTGAGCCGTGCTGCTGGAAGAGCCGAGATCATAGACCATGCCCGCGCGCTG	721

```

Db      ||| 601 ACTGCACATGAGCGCGTGCCTCGTGAAGGCCACAGATCATAGGCATCCCGCGCGTG 660
        ||| 722 GCGGTATAGAGAGGGGACTTTGCGCCCTCGCAAAAAGGGCAAGATGGGACTTCGCTCAAC 781
        ||| 661 GCGGTATAGAGAGGGGACTTTGCGCCCTCGCAAAAAGGGCAAGATGGGACTTCGCTCAAC 720
        ||| 782 GCGCATCTATATGAGCCCTGGAGACAGATAGAGCTCGGAGCAAGAGGCTGTAAGCA 841
        ||| 721 GCGCATCTATATGAGCCCTGGAGACAGATAGAGCTCGGAGCAAGAGGCTGTAAGCA 780
        ||| 842 CCGATGGAATTTTCACTTGGCTGTTTGCCTCAATCCCATCTTTGAAGAGGACTATCA 901
        ||| 781 CCGATGGAATTTTCACTTGGCTGTTTGCCTCAATCCCATCTTTGAAGAGGACTATCA 840
        ||| 902 GAGAGCATGAGAGAGAGCTGGGCGAGAGGCTTCAGAGCCCTCAGCCCGGCACTTTGCC 961
        ||| 841 GAGAGCATGAGAGAGAGCTGGGCGAGAGGCTTCAGAGCCCTCAGCCCGGCACTTTGCC 900
        ||| 962 ATCTCTAATGCGGAGAGACCGACTTCTACGCGCATGAATTAATACACATCCAGTTCCG 1021
        ||| 901 ATCTCTAATGCGGAGAGACCGACTTCTACGCGCATGAATTAATACACATCCAGTTCCG 960
        ||| 1022 CCGCATCTATGAGAGGCTTCGCTCCCGAGACGACTATTCGCGCGCATCATGAGACGAG 1081
        ||| 961 CCGCATCTATGAGAGGCTTCGCTCCCGAGACGACTATTCGCGCGCATCATGAGACGAG 1020
        ||| 1082 GAGATTAAGAGAGCGAGCGCCGTTGGCGAGAGAGCGGCTCGGCTGCGTCCGCTCCG 1141
        ||| 1021 GAGATTAAGAGAGCGAGCGCCGTTGGCGAGAGAGCGGCTCGGCTGCGTCCGCTCCG 1080
        ||| 1142 CCGCATCTATGAGAGGCTTCGCTCCCGAGACGACTATTCGCGCGCATCATGAGACGAG 1201
        ||| 1081 CCGCATCTATGAGAGGCTTCGCTCCCGAGACGACTATTCGCGCGCATCATGAGACGAG 1140
        ||| 1202 ATCAACGAGAGAGCGAGCGCCGTTGGCGAGAGAGCGGCTCGGCTGCGTCCGCTCCG 1261
        ||| 1141 ATCAACGAGAGAGCGAGCGCCGTTGGCGAGAGAGCGGCTCGGCTGCGTCCGCTCCG 1200
        ||| 1262 AACGACCCCTTCGCGATCCGCTGCTTGAATCTCGACTGCACTGATTCGATTCGAGGCCAT 1321
        ||| 1201 AACGACCCCTTCGCGATCCGCTGCTTGAATCTCGACTGCACTGATTCGAGGCCAT 1260
        ||| 1322 ACCGAGAGCGGCGTGTCTGTAAGGGGTAATTTGGCTGCGTTCGATTAACCTTGGA 1381
        ||| 1261 ACCGAGAGCGGCGTGTCTGTAAGGGGTAATTTGGCTGCGTTCGATTAACCTTGGA 1320
        ||| 1382 TGTCTAGATGGCTAGCGAGCCGAGATTGGGGTCACTGCTACAGACTACACACCTTCAG 1441
        ||| 1321 TGTCTAGATGGCTAGCGAGCCGAGATTGGGGTCACTGCTACAGACTACACACCTTCAG 1380
        ||| 1442 CGCAGCGCCAGAAAGTCTGCTGCTCAAGAGCAATGTTGGCGCCGCGCAGAGGGTT 1501
        ||| 1381 CGCAGCGCCAGAAAGTCTGCTGCTCAAGAGCAATGTTGGCGCCGCGCAGAGGGTT 1440
        ||| 1502 AAAGTGGGGGATTA 1516
        ||| 1441 AAAGTGGGGGATTA 1455

```

```

KW      Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW      culture condition; environmental stress; spore morphogenesis;
KW      metabolic pathway engineering; catabolic pathway engineering; ss.
OS      Hypocrea jecorina.
PN      MO200056762-A2.
PD      28-SEP-2000.
PF      22-MAR-2000; 2000WO-US007781.
PR      22-MAR-1999; 99US-00273623.
PX      (NOVO ) NOVO NORDISK BIOTECH INC.
PA      (NOVO ) NOVO NORDISK AS.
PI      Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
PI      WPI; 2000-594572/56.
PT      Monitoring differential expression of genes in filamentous fungal cells
PT      uses fluorescence-labeled nucleic acids isolated from the cells and a
PT      substrate of expressed sequence tags.
PS      Claim 89; Page 3012; 3161dp; English.
XX      The present invention describes a method for monitoring differential
XX      expression of genes in a first filamentous fungal (FF) cell relative to
XX      expression of the same genes in one or more second filamentous fungal
XX      cells. The method uses fluorescence-labeled nucleic acids isolated from
XX      the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX      are used in the methods for monitoring differential expression of genes
XX      in a first filamentous fungal (FF) cell relative to expression of the
XX      same genes in one or more second filamentous fungal cells. Monitoring the
XX      global expression of genes from FF cells allows the production potential
XX      of the microorganisms to be improved. New genes may be discovered,
XX      possible functions of unknown open reading frames can be identified and
XX      gene copy number variation and stability can be monitored. The expression
XX      of genes can be used to study how FF cells adapt to changes in culture
XX      conditions, environmental stress, spore morphogenesis, recombination,
XX      metabolic or catabolic pathway engineering. Using ESTs provides several
XX      advantages over genomic or random cDNA clones including elimination of
XX      redundancy as one spot on an array equals one gene or open reading frame,
XX      and organization of the microarrays based on function of the gene
XX      products to facilitate analysis of the results. AA07478 to AA07478
XX      represents ESTs from Fusarium venenatum; AA07478 to AA07478 represents
XX      ESTs from Aspergillus niger; AA07478 to AA07478 represents ESTs from
XX      Aspergillus oryzae; and AA07478 to AA07478 represents ESTs from
XX      Trichoderma reesei, which are all specifically claimed in the present
XX      invention. (Updated on 15-SEP-2003 to standardise OS field)
SQ      Sequence 735 BP; 168 A; 210 C; 205 G; 148 T; 0 U; 4 Other;
Query Match 30.0%; Score 598; DB 3; Length 735;
Best Local Similarity 97.0%; Pred. No. 1e-137;
Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;

```

QY 559 GATCACCATCAAGMCCCTGATTGAGCCATCTATGATATGCAACCGGACGAACGC 618
 Db 241 GATCACCATCAAGMCCCTGATTGAGCCATCTATGATATGCAACCGGACGAACGC 300
 QY 619 CCGGGGAGAGAGCATTAACAAGCACTCCAGGAGGAGCACTGCCACTGAGCCGTG 678
 Db 301 CCGGGGAGAGAGCATTAACAAGCACTCCAGGAGGAGCACTGCCACTGAGCCGTG 360
 QY 679 GCTGCTGGAAGAGCCCATGATCATGAGCCATGCCCGCGCGTGGCCGTCTACAGCAGGA 738
 Db 361 GCTGCTGGAAGAGCCCATGATCATGAGCCATGCCCGCGCGTGGCCGTCTACAGCAGGA 420
 QY 739 CTTTCCGCCCCCGCAAAAGGCGGAGATGCGCATGCTGCTCAACGGGAGTACTATGAGCC 798
 Db 421 CTTTCCGCCCCCGCAAAAGGCGGAGATGCGCATGCTGCTCAACGGGAGTACTATGAGCC 480
 QY 799 CTGGGACAGCATGAGCTCTGGGCAAGAGAGGCTGTAGCGACGATGGAATTTCAAT 858
 Db 481 CTGGGACAGCATGAGCTCTGGGCAAGAGAGGCTGTAGCGACGATGGAATTTCAAT 540
 QY 859 TGCTGCTTTGCCAATCCATCTTTTGAAGAGAGCTA-TCCAGAGCATGAAGAAC 917
 Db 541 TGCTGCTTTGCCAATCCATCTTTTGAAGAGAGCTAATTCNAAGAAAGCATGAAGAAC 600
 QY 918 AGCT--GGGAGAGAGCTTCAGCCCTCAC-TCCGGGAGCTTT--GCCATCCCTCAATGC 972
 Db 601 AACTTGGGGGAGAGGCTTTCACCTTACCTTCCCGGAGACTTTTGCATCTTNAATGC 660
 QY 973 CGGAGAGAGCC 982
 Db 661 CGGAGAGAAC 670
 RESULT 4
 AAF14882
 ID AAF14882 standard; cDNA; 1933 BP.
 AC AAF14882;
 XX
 DT 15-SEP-2003 (revised)
 DT 13-MAR-2001 (first entry)
 XX
 DE Trichoderma reesei EST SEQ ID NO:7405.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 OS Hypocrea jecorina.
 OS
 PN W0200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000MO-US007781.
 XX
 PR 22-MAR-1999; 99US-00273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 PI Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
 DR WPI; 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags.
 XX
 PS Claim 89; Page 2989; 3161p; English.

XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring the
 CC global expression of genes from FF cells allows the production potential
 CC of the microorganisms to be improved. New genes may be discovered,
 CC possible functions of unknown open reading frames can be identified and
 CC gene copy number variation and stability can be monitored. The expression
 CC of genes can be used to study how FF cells adapt to changes in culture
 CC conditions, environmental stress, spore morphogenesis, recombination,
 CC metabolic or catabolic pathway engineering. Using ESTs provides several
 CC advantages over genomic or random cDNA clones including elimination of
 CC redundancy as one spot on an array equals one gene or open reading frame,
 CC and organization of the microarrays based on function of the gene
 CC products to facilitate analysis of the results. AF07478 to AAF11247
 CC represents ESTs from *Fusarium venenatum*; AAF1248 to AAF11853 represents
 CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from
 CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from
 CC *Trichoderma reesei*, which are all specifically claimed in the present
 CC invention. (Updated on 15-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 1933 BP; 467 A; 536 C; 568 G; 343 T; 0 U; 19 Other:
 Query Match 13.4%; Score 266.6; DB 3; Length 1933;
 Best Local Similarity 56.2%; Pred. No. 1.5e-55;
 Matches 770; Conservative 1; Mismatches 540; Indels 59; Gaps 13;
 QY 81 TGCCCAACGACTTGAATGGGCTTGCCAAAGCCGCTCAACGATGAGAGGCGCTCA 140
 Db 250 TGCCCAAGACTTTCAGTGGGAGTTCGCCACGCTGCTACAGATGAGAGGCGCTCG 309
 QY 141 AAGAAGTGCGCGCGCGCGCTCCAT--CTGGACACAGTACTGC-----CACTGAGACCA 193
 Db 310 ACCAGAGAGCG 369
 QY 194 TCGCGCACCAAGCG 252
 Db 370 TCGCGCACGCGCTNCGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 429
 QY 253 GGAATTGATCTCTTGACCAAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
 Db 430 GGAATTGCGCT 489
 QY 313 GCGGATCATTCCTCG 372
 Db 490 GCGGATCATTCCTCG 549
 QY 373 CAGCAACTGATTGACGCGCTGTGAGGCGGGGTATCAGCGCTTGGGTGACTTTGATCA 432
 Db 550 CGTCAAGTTGCTGAGAGACTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
 QY 433 CTGGGATCTGCTCAAGCGCTTACGATCTGTATGAGCGTGTCTCAAGTGAAGAGT 492
 Db 610 CTGGGATCTGCTCAAGCGCTTACGATCTGTATGAGCGTGTCTCAAGTGAAGAGT 669
 QY 493 CCAAGCTGACTTGAAGCGGATGAGAGTGTGCTTGAAGCTTTTGGGAGCGAGTCA 552
 Db 670 CCGGCTGACTTGAAGCGGATGAGAGTGTGCTTGAAGCTTTTGGGAGCGAGTCA 728
 QY 553 GAACTGATCAACATCAAGMCCCTGATTGAGCCATCTATGATATGCAACCGGACG 612
 Db 729 CAACCTGATCAACATCAAGMCCCTGATTGAGCCATCTATGATATGCAACCGGACG 788
 QY 613 CAACGCCCCGGGAGAGAGCATTAACAAGCACTCCAGGAGGAGCAACTGCCACTGA 672
 Db 789 CTTGCCCCCGGCGGAGAGCA-----CTTCGGA 818
 QY 673 GCGTGGCTGCTGGAAGAGCCAGATCATGAGCATGCGCGCGCGCGCGCTTACAG 732

Db 819 GCGGTGACCGTGGCCCAACAATCTCTGCTGACCCAGCGCGCCCTCAAGCGTACC 878
 QY 733 CAGGAGCTT---TCGCCCCCTCGCAAAAGGCGAGATGGGATCGCTCAACGGCGACTA 789
 Db 879 CGAGACTTCAAGCCCCCGACGGCGGCGAGCGGCAATGGGATCGCTCTCAACGGCGACTT 938
 QY 790 CTATGAGCCCTTGGAGACGACATGAGCCTCGGCGACAGAGGCTGCTGAGCGAGCGATGGA 849
 Db 939 CACTTACCCCTGGAGACCGCCCGGACCGGCGGACAGAGGCGGCGGCGGCGCTCGA 998
 QY 850 ATTTCATATGGCTGCTGTTTGGCAATCCCATCTTTTGAAGAAGACTATCCAGAGACAT 909
 Db 999 GTTCTTCAACGCGCTGTTTGGCGGATCCCATCTACTTGGGCGACTACCGCGGCTCGA--TGC 1057
 QY 910 GAAGAGACGCTGGGCGGAGAGCTTCCAGCCTCACTCCGCGGACTTTGGCATCTCA 969
 Db 1058 GCAAGCAAGCTGGGCGGACCGGCTGCGGACCTTTACCGCCGAGGA--AGCGCGCCTCGT 1115
 QY 970 TGGCGAGAGACCGACTTCTACGCGATGATTAATCAATCCAGTTGCGCGGCGACT 1029
 Db 1116 CCAGCGCTCCACGACTTTTACGGGATGACCACTACGTCACATACATCCGCGACCGA 1175
 QY 1030 AGACGCTCCGCTCCCGGAGAGAGACTATCTGGGCGGCGCATCATGAGACGAGAAATA 1089
 Db 1176 AGCTCG--CCGCTTGGCGGAGACGACCGTGGCGAAGCTGCACTGCTTCCACCAACA 1233
 QY 1090 GAGCGGACGCGCCGCTGGCGAGAGAGCGGCTCGCTGCGTGGCTCT--GGCCGAGC 1147
 Db 1234 GCAAGGCAATGTCATCGGCGCCGAGGA--CGAGTCCCTCGGCTGGGCGCCCTTGGCGCG 1292
 QY 1148 ATGTTCCGGAAGCATCTGCGCGGCTGTACGCGCTGTACGCGCAAGCCCATCTACATCAC 1207
 Db 1293 GNTTTCGCCGANTTCTMGSTGTGATTCAGCAAGANGACMGSTACCCGCCATTNTACGT 1352
 QY 1208 GAGAACGATGCGCGTGCCTTGAGAGAGAAATGACGTGC-----GAGAGAGCGCT 1260
 Db 1353 GACGGANACGAGNACGAAAGCATTAAGGCGAGAGGACTTGGCAAGAGAAAGAAATNT 1412
 QY 1261 CAACGACCCCTTCCGCAATCCGCTTGTGACTGCACTTGGCATGCTGAT--TTCCAGAGCC 1318
 Db 1413 CGAAGTGACTTCAAGGTCAATTAACAAGTACATCCGTCATGGGTATCCGCGCG 1472
 QY 1319 ATTTACCGAGAGCGGCTGCTCTCAAGGAGTCTTGGTGGGCGTGGCTGATTAATCTTG 1378
 Db 1473 TGGAGCTGAGCGGGGTCAAAAGCTCAAGGGTACTTGGCTGCTGCTCATGGAACAATT 1532
 QY 1379 GAATGTCAGATGCTGACGAGACCGAGATTGCGGCTGACGTTCAAGACTTA 1428
 Db 1533 GAGGGGGCGGAGCGCTACGTAAGAAAGTTTGGGGTTACTATGGGGAATTA 1582
 RESULT 5
 ADA70881
 ID ADA70881 standard; DNA; 1404 BP.
 AC ADA70881;
 AC 20-NOV-2003 (first entry)
 DT
 XX
 DE Rice gene, SEQ ID 4204.
 KW
 XX Plant; bacterial infection; fungal infection; viral infection; rice;
 KW gene; ds.
 XX
 OS Oryza sativa.
 XX
 PN MO2003000898-A1.
 XX
 PD 03-JAN-2003.
 XX
 PF 22-JUN-2001; 2001WO-1B001105.
 XX

PR 22-JUN-2001; 2001WO-1B001105.
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 XX WPI, 2003-175290/17.
 DR
 PT Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.
 PS
 XX Claim 6; SEQ ID NO 4204; 899bp; English.

The present invention relates to a method (M1) for identifying genes
 involved in plant resistance or response to pathogenic infection. M1
 comprises identifying a gene whose expression is significantly altered in
 the incompatible interaction of plant gene expression relative to
 expression of the gene in an uninfected plant, in a mutant plant that
 does not express a gene associated with response to pathogenic infection,
 or in a corresponding incompatible or compatible interaction. (M1) is
 useful for conferring resistance to resistance or tolerance to a plant to
 bacterial, fungal or viral infection. The present sequence was used to
 illustrate the invention.

Sequence 1404 BP; 356 A; 385 C; 396 G; 266 T; 0 U; 1 Other;

Query Match 12.4%; Score 246; DB 7; Length 1404;

Best Local Similarity 54.5%; Pred. No. 1.7e-50; Matches 725; Conservative 0; Mismatches 565; Indels 40; Gaps 10;

QY 123 AGATCGAAGCGCGCTCAAGAAAGTGGCGCGCGCCGCTCATCTGAGACACTGTCGCC 182
 Db 86 AGTACGAGGAGCTGTGAGAGAGAGCGAGAGGCGAGACATCTGGGACACTTTGGCGC 145
 QY 183 ACTTGAGCCATCGCGCACCAACGCGCGCAAGCGCGATGTGGCTTGGCATCTACCAACC 242
 Db 146 ACACCTTTGGAAAGATCACCGACTTCAGCAATCTGTATGTTGCAAGTTGATCAGTACCAACC 205
 QY 243 GCTACGATGAGACTTGTATCTCTTACCAAGTACAGGCGGAAAGGCGCTTCTCTCT 302
 Db 206 GTTTCGAGAGAGATACACTCATGCGCAACATGGGAGATGCTTATCGGTTCTCGA 265
 QY 303 TGTGATGATCGCGGATATTCCTCGCGCGCAAGCTGATATCCCGTCAACGAGAGGAAA 362
 Db 266 TAGCATGCTCAAGATATACCAAAATGTG-----TTGGTCAAGTCAATCAAGCTGTA 319
 QY 363 TTAGATTTTACAGCAAACTGATTGACGCGCTGTTGAGCGGGGTATCAAGCTTGGGTGA 422
 Db 320 TCGACACTACACAACTGATGATGATCTTACGCAAAAGAAATTCAGCATATATGTA 379
 QY 423 CTGTGTACCACTGGATCTGCTCAGCGGCTTTCAGATCGGTATGAGAGGCTGCTCAACG 482
 Db 380 CACTTACCACTGGAGCTTCCCGAGGCTTTGAAGACAACTAACAGGCGTGGCTTGAACA 439
 QY 483 TGAAGAGTCCAGCTGACTTTGAGCGGTATGCGAGGTGTGCTTTGAAGCTTTTGGGG 542
 Db 440 GG---CAGATAGTGGAGATTTGCGGGGTACGCGGAGACGCTGCTTCAAGGAGTTTGGGG 496
 QY 543 ACCGATCCAGACTGATCACTCAATCAAGCAAGWCCCTGGATTCAAGGCATATGATGATG 602
 Db 497 ACAGGTGAAACACTGATGATCAAGCTCAACGAGCGCACACGCTGACCAATCAAGGCTTACG 556
 QY 603 CCACCGGAGAGACGCGCGCGGAGAGCAGACTTA-----ACAAGACTTCAACGAGG 656
 Db 557 ACGAGGCTCAGGCGCCCGCGCGCTCTCTGCTGCTCTCACTCAATCAAGGCGCG 616
 QY 657 GCAACATGCACTGAGCGGTGCTGCTGGAAGAGCCCAATCATGAGCCATGCGCGCG 716
 Db 617 GCAATCCGGGACGAGCCTTACGCTGCGCCACCACTTATCTCTGCCACCGCGCGCG 676

QY 717 CCGTGGCGCTTACAGAGGAGGACTTTCGCCCTCGCAAAAGGCGCAGATCGCATCTCGC 776
 DB 677 CCGGACCATCTACAGAGCAAAATACAAAGCGACGAGAAACGGGACACTTGGGATACCT 736
 QY 777 TCACGGCGACTACTATGAGCCCTGGGACAGCAATGAGCCTCGGACAAAGAGCTGCTG 836
 DB 737 TCGACGTGATGTGTGAGCCGACGATGTCACACAGAT---CGACATGAGAGCGGCGCA 793
 QY 837 AGGACGCGATGAAATTTACATGCGCTGCTGCTTCCATCTCCATCTTTTGAAGAGACT 896
 DB 794 AGAGAGCGAGAGTTTACATGAGATGTTTGTGATCGCTTCTT---CGGCACT 850
 QY 897 ATCAGAGAGCATGAAAGAACAGCTGGGCGAGAGCTTCCAGCCCTCACTCCGCGGACT 956
 DB 851 ACCCGGAGCATGAGGCGGAGAGCTGGGAGAGAGCTCCGAGTTTCAACGCGGATAGG 910
 QY 957 TTGCCATCTCAATGCCGAGAGACCGACTTCTAAGGATGAAATTTACTACATCCAGT 1016
 DB 911 CCGCGCTGCTCAAG---GGGCGCTGATTTCTGCGCATTAACCACTACACCACTACT 967
 QY 1017 TCGCGCGCCACCTAGAGGCTCCGCTCCCGAGA-----CGGACTATCTCGGCGCATCA 1071
 DB 968 ACAGAGGCGACAAACAAACCAACATCATCGGAGATTTCTCAACACACCTTGGCAGACA 1027
 QY 1072 TGAGCAGCAGAGAAAT-----AAGAGCGAGAGCCCGTTGGCGAGAGAGCGGCTCG 1124
 DB 1028 CCGGACCGCTCAAGCTCCCATTTCAAGAAATGGGAAGCCAAATTTGAGATAGGCAATTCGA 1087
 QY 1125 CTTGGCTGCGCTCTGCGGAGACATGTTCCGGAAGCATCTCGCCGGGTGTACGCGCTGT 1184
 DB 1088 TATGCTGTGATTTGTGCTCCCGAGAGATGAGAGCTGATGAATGATCAAGGAAGGT 1147
 QY 1185 ---ACGGCAAGCCCATTAATCATACCGAGAACGATCCCGCTCGAGAGAGAAACA 1241
 DB 1148 ACAAGAGCCGACAGTGTATCATCTGAAACCGGATGATGACAGCAACACCGCTTCA 1207
 QY 1242 TGAGCTGAGAGAGCGGCTCAACGACCCCTTCCGATCCGATCTTGAATCGCATTTGG 1301
 DB 1208 TTTTCATCAAGAGCGCCCTCAAGAGCAGCAAGAGATCAATATCAAAATGACTACCTCA 1267
 QY 1302 ACTGATTTCCAAAGCCATTACCGAGAGCGGCTGCTGTCAAAGGCTACTTTGCGTGG 1361
 DB 1268 CCAATCTGCTGCTTCCATCAAGAGAGCGGCTGCGATACGATGCTACTTCCGCTGTG 1327
 QY 1362 CGTTGCTGCTTAATCTTGAAT-TGTCAGATGCTGCTAGGAGCCAGATTTCCGCTGCTG 1420
 DB 1328 CTCTGCTGCAACTGGGAATGGGCGGCGGATACCTCTGAGATTCGGGCTTACTTTC 1387
 QY 1421 ACAGACTACA 1430
 DB 1388 GTGGAATACA 1397

RESULT 6
 ADA69524
 ID ADA69524 standard; DNA; 1503 BP.

AC ADA69524;
 XX 20-NOV-2003 (first entry)
 DT XX
 DE Rice gene, SEQ ID 2847.
 KW plant; bacterial infection; fungal infection; viral infection; rice;
 XX gene; ds.
 OS Oryza sativa.
 XX
 PN MO2003000898-A1.
 XX
 PD 03-JAN-2003.
 XX
 PF 22-JUN-2001; 2001MO-IB001105.

XX
 PR 22-JUN-2001; 2001MO-IB001105.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
 PI Katsagiri F, Quan S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;
 XX
 DR WPI; 2003-175290/17.
 XX
 PT Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.
 XX
 PS Claim 6; SEQ ID NO 2847; 899bp; English.
 XX
 CC The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.
 XX
 SQ Sequence 1503 BP; 380 A; 420 C; 379 G; 324 T; 0 U; 0 Other;
 Query Match 10.8%; Score 214.6; DB 7; Length 1503;
 Best Local Similarity 51.0%; Pred. No. 1e-42;
 Matches 720; Conservative 1; Mismatches 660; Indels 30; Gaps 8;

QY 81 TGTCCAAAGCACTTGAATGAGGCTTTCGCAAGCGCCCTCAACGATGAAAGCGCGCTCA 140
 DB 95 TCCCAAGAGCACTTCACTTTCGCGACAGAGTTCAACAGCTTATCAGTATGAGGCGCTTCA 154
 QY 141 AATAAGTGGCGCGCGCGCGCTCCATCTGAGAGACAGTATCCGACCTGAGACCATCGGCA 200
 DB 155 AATAAGGCGCGCGCGCGCTTCAAGGCTGAGAGCACTTCACTCACTTCCAGTAAATTT 214
 QY 201 CCAAGCGCGCGCGCGAGATGATGCTTTCGATCAACGCGCTAGAGAGACTTTG 260
 DB 215 TAAACATATGATTAACGCGAGATGAGCAATGATCTTATCAACGATCAAGAGAGATGTA 274
 QY 261 ATCTTTCGCAAGTACGCGCGCAAGCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 320
 DB 275 GCTCTGTAAGAGACATGATGATGATGCTTTCGCTTCTCATTTGCTGAGACGAGATTC 334
 QY 321 TTCCCTCGCGCGAGCTGATCCGCTCAACGAGAGGAAATTTAGTTTACAGCAAC 380
 DB 335 TGCCAATGATCTTGAATGAGAGAAATAAACAAGAGGATTTGCTTTCACACGCT 394
 QY 381 TGAATGACCGCTTGTGAGGCGGAGTATCAAGCGCTTGGGATCTTTGATCACTGGGATC 440
 DB 395 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 454
 QY 441 TGCCTGAGCGCTTTCAGATCGCTATGAGAGCTGCTCAACGTAAGAGAGTCAAGCTGG 500
 DB 455 CCCCCTTGGCTTGAAGAGCAATTAACGAGAGATTTCTCA---GTGAAGACATAGTGAAG 511
 QY 501 ACTTTGAGCGGTATGAGAGTGTGCTTTCGTAAGCTTTTGGGAGCCGAGTCAAGACTGGA 560
 DB 512 AATACGTGACTTCGCGAGAGTGTGCTTTCGAGATTCGCGAGCGGCTGTCATTAATCTGGA 571
 QY 561 TCACATCAAGAACCTGATTCAGAGCATATGATGATATGCAACGCGAGCAAGCGCC 620
 DB 572 CCAATTTAATGAGCATTCATATAGCGCTTACGCGCTACGCGAGGCGGTGTTTGGCG 631
 QY 621 CCGGAGAGAGCAACA-----TTAACAGCACTTCACGAGGCAACATCTGCCACTGAGC 674
 DB 632 CCGGAGAGAGTCTCTCATATGTTTCCAAAGTCATGCGGCTGCTGCTGCTGCTGCTGCTG 691

QY 675 CGTGGCTGCTGAGGAGCCCAATCATGAGCCATGCGCCGCGTCTACAGCA 734
 DB 692 CCACTCTGTGGACACCAATCCATCTTCCACGCTGCGCTTCAAGTAAACGCA 751
 QY 735 GGGACTTTGGCCCTTGGCAAAAGGCGAGATCGGCACTCTGCAACGGCACTACTATG 794
 DB 752 CCAAGTACCAAGCAACAGAAAGGACAGATCGGATGGTGGTGTCAACCCACTGGTTCG 811
 QY 795 AGCCCTGGAGACAGATGAGCTCGGGACAAGAGAGCTGTGAGGACGAGATGGAATTC 854
 DB 812 TGCCTGACAGCACTCCAGCT---GACGCTGGCGCTGTGCAACGAGACCTTGAAGCTTCA 868
 QY 855 ACATTGCTGATTTGCCAATCCCATCTTGAAGAAGAGATCATCAGAGAGATGAAGA 914
 DB 869 TCTATGGGTGTTCATGAGACCTTA---TCGTGATGATGATCAACCAAGCAACATGAGAG 925
 QY 915 AGCAGCTGGGCGAGAGGCTTCCAGCCCTGACTCCCGGAGCTTTGCAATCTCAATGCGG 974
 DB 926 GTTGGCTGGCAATCGGCTGCGAGATTCAAGCTGAAACAGTCGGCGATGGTGAAGGGCT 985
 QY 975 GAGAGACCGACTTCTACGCGATGAATTAACAATCCAGTTCGCGGCGCACTAGACG 1034
 DB 986 CCTA---GCACTTCACTCGAGTAACTTACTACACCACTTACGCTTAAAGTAAACCGC 1042
 QY 1035 GTCCGCTCCCGA-----GACGACTATCTCGCGCGCATCATGAGCAGCAGAGATA 1088
 DB 1043 CGCCTTACTCCAAAGCATATCTTACGACTGACCTGACACCGCGCAACACACCGGCTTCC 1102
 QY 1089 AGGACGGAGAGCGCGTGGGCGAGAGAGAGCGGCGCTGCGCTGCGCTCTGCGCGGACA 1148
 DB 1103 GTAAATGGCAAAACCATCGGTCCACAGGAATTTACACCATCTTCTTCAACTAACCTCCAG 1162
 QY 1149 TGTTCGGAAGATCTCGCCCGGAGTGAACGCTGTGACGAGAGCCCATTTACATCAACG 1208
 DB 1163 GTCTCGTAGAGTCTCTCTTACACCAAGAGAGATTAACAACCCGA---CAATTAATG 1219
 QY 1209 AGAAGGATGCGCGCTGCGAGAGAGAAATGATGATGATGATGATGATGATGATGATG 1268
 DB 1220 TTACAGAAACCGGACATCATGAGGATTAACAACAGACACTGCCAGAGGCGCTCAAGATG 1279
 QY 1269 CTTTCGCGATCGGATCTTGAATCGCATGCGATTTTCAAGGCGATTAACCAAG 1338
 DB 1280 GACACAGATGAGATTCATCAACAGACCTGCGATTCGTAACACAGCGCAT---CAAGA 1336
 QY 1329 ACGGCTGCTGCTCAAGGGTACTTTGCGTGGGCGTGTCTGATTAATCTTGATGATGTCAG 1388
 DB 1337 ATGGGTGTAAGCTGAAGGGTACTTCAAGTGAATTCATGATGATGATGATGATGATG 1396
 QY 1389 ATGGCTACGACCAAGATTTGGCGTCAAGTTCACAGATCAACCACTTCAAGCGACGC 1448
 DB 1397 ACGGCTACCTTGAAGGTTGGCTCTCATACGTCGACCGGACAGCTCAAGGCTTACC 1456
 QY 1449 CCAAGAAGTCTGCGCTGCTCTCAAGGACAT 1479
 DB 1457 GCAAGAGTCAAGTCACTGATCAAGACTT 1487

RESULT 7
 ABZ53130
 ID ABZ53130 standard; cdna; 575 BP.

XX ABZ53130;

DT 28-MAR-2003 (first entry)

DE Aspergillus oryzae polynucleotide seq ID NO 2243.

XX Aspergillus oryzae; fermentation; fungus; industrial; BST;

KW expressed sequence tag; gene; ss.

XX Aspergillus oryzae.

XX

PN W0200279476-A1.
 XX 10-OCT-2002.
 PD 22-MAR-2002; 2002WO-1B000890.
 XX 30-MAR-2001; 2001JP-00098371.
 PF (NABD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX (NABE-) NAT RES INST BREWING.
 PA (NORQ) NAT FOOD RES INST MIN AGRIC.
 PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiiuchi H;
 PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;
 DR WPI; 2003-046817/04.
 XX
 XX
 PT Detection of expression of specific *Aspergillus* genes for monitoring the
 PT fermentation and growth conditions of the fungus, using DNA probes.
 PS Claim 1; SEQ ID NO 2243; 48bp + Sequence Listing; Japanese.
 CC The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridizing
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus,
 CC especially of *Aspergillus oryzae* which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination. Note: The
 CC sequence data for this patent did not form part of the printed
 CC at fwp.wipo.int/pat/published_sequences
 CC
 XX
 SQ Sequence 575 BP; 143 A; 154 C; 134 G; 144 T; 0 U; 0 Other;
 Query Match 10.3%; Score 205.4; DB 7; Length 575;
 Best Local Similarity 63.0%; Pred. No. 1.3e-40;
 Matches 317; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
 QY 78 CTCTGCCCAACGACTTGAATGGGCTTGGCAACGCGCGCTTACAGATGAAGGCGCG 137
 DB 71 CTCTCCACCGGACTTCTATGGGATTCGCAACGCGAGTATACAGATTTGAAGGCGCG 130
 QY 138 TCAAGAAGTGGCGCGCGCGCGCTGCATCTGGAGACAGTACAGCCTGAGGACATGCG 197
 DB 131 TAAATAGAGACGAGCAGGTCCATCTGATGATCAATTGGCAAGATCCCGGTAAAA 190
 QY 198 GCAACACGCGCGCAACGCGAGTGTGCTTGCATCACTACACCGCTTACATGAGACT 257
 DB 191 TTGCTGAGAGCGCGCAATGCGATGTGGCTTGGACATCAATACATGACACAGAGATA 250
 QY 258 TTGATCTCTTGAACAGTACGCGCGCAAGGCTTACCGCTTCTCTTGTGTGTCGCGA 317
 DB 251 TTGCTTGTCTCAAGGCTTGGCTGCGCAAGGCAATACGTTTCTCACTCTTGTGTCGCGA 310
 QY 318 TCAATTCCTCGCGCGCGAGCTGATCCGCTCAAGAGAGGAGTATGATTTACAGA 377
 DB 311 TTATTTCACTCGGTGGCGCAACGACCCCATATATGAAAGGCTTGCATATATATATA 370
 QY 378 AACTGATGACGCGCTGTTGAGGCGGGTATCACCGCTTGGGTGACTTTGATCACTGGG 437
 DB 371 AGTTGCTGACGACCTGCACTGCTGCTGATCACTCTCTGTTACTTTGTTCACTGGG 430
 QY 438 ATCTGCTTACGCGCTTACAGATCGCTATGAGAGCTGCTCAACCTGGAAGAGTCCAGC 497
 DB 431 ATCTTCCGATAGGTTCACAAACGCTACGCGGCTCTCTTCAATAAGAGAGTTCGTCG 490
 QY 498 TGAAGCTTGAAGCGGATGAGAGTGTGCTTTGAAGCTTTTGGGAGACGAGTCCAGACT 557
 DB 491 CAGACTTGGCCCACTACGACGACGATGATTTTCAAGGCTTTGGGCTCGAAGATTAAGCA 550

Dh 1341 AGTGAAGATGGTGTGACGTAAAGAGATATTACGATGCTGATTGATGACAATTTTGA 1400
Qy 1381 ATGTCAGATGGCTAACGACCCCAATTCCG 1410
Dh 1401 ATGGAGCATGATGATCACTGCAAGATTTCG 1430

RESULT 9

ADA67833
ID ADA67833 standard; DNA; 1734 BP.

ADA67833;

20-NOV-2003 (first entry)

Arabidopsis thaliana gene, SEQ ID 71.

Plant; bacterial infection; fungal infection; viral infection; ds.

Arabidopsis thaliana.

WO2003000898-A1.

03-JAN-2003.

22-JUN-2001; 2001WO-1B001105.

22-JUN-2001; 2001WO-1B001105.

(SYN) SYNGENTA PARTICIPATIONS AG.

Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
P1 Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

WPI; 2003-175290/17.

Identifying at least one gene involved in plant resistance or response to
pathogenic infection for conferring resistance or tolerance to a plant to
bacterial, fungal or viral infection by determining or detecting plant
gene expression.

Claim 6; SEQ ID NO 71; 899bp; English.

The present invention relates to a method (M1) for identifying genes
involved in plant resistance or response to pathogenic infection. M1
comprises identifying a gene whose expression is significantly altered in
the incompatible interaction of plant gene expression relative to
expression of the gene in an uninfected plant, in a mutant plant that
does not express a gene associated with response to pathogenic infection,
or in a corresponding incompatible or compatible interaction. (M1) is
useful for conferring resistance or tolerance to a plant to
bacterial, fungal or viral infection. The present sequence was used to
illustrate the invention.

Sequence 1734 BP; 556 A; 316 C; 385 G; 477 T; 0 U; 0 Other;

Query Match 9.2%; Score 182.8; DB 7; Length 1734;

Best Local Similarity 49.2%; Pred. No. 1.e-35;

Matches 664; Conservative 1; Mismatches 656; Indels 27; Gaps 6;

Qy 79 TCTGCCAAGCATTTGAAATGGGCTTCGCAACGCGCCGCTTACGATCGAAGGCGCGT 138
Dh 90 TTTTCTGATGATTTTCATCTTTGGACAGCTGCTCGGCTTCAAGTACGAAGTGCAC 149
Qy 139 CAAGAAGTGGCCCGGCGCGCTTCATCTGGGACACGTACTCCACCTTGGAGCCATCGG 198
Dh 150 AAGTAAGGTGGCAAGTCTCCCAACTAATAGGATCACTTCAGCTTCACGTATCCGAAG 209
Qy 199 CACCAAGCGCGCCACGCGGATGGCTTGGCATCACTACCGCTACGATGAGGACTT 258
Dh 210 GACCAAAATGCAATATGACGATGACGAATTAATTTTATCATCTGTACAAAGATGACAT 269
Qy 259 TGATCTTTGACCAAGTACGCGCAAGGCTTACGCTTCTCCTTGTCTGTGCGGAT 318

Dh 270 AAAATTGATGAAGAGCTAAACATGACGCTTTCCGATTTTCAATCTCGGGTCAAAATT 329
Qy 319 CATTCCTCCCTGGCGGACGCTGATCCCTTCACACGAGAGGAAATGATTTTACAGCA 378
Dh 330 AATACCAATGAAAGCTTAAAGATGAGTAACAAAGAGGTACATTTCTACAAAGCA 389
Qy 379 ACTGATTGACGCTCTTGTGAGGCGGGTATACGCTTGGGTGATCTTTGTAACAATGGGA 438
Dh 390 TCTCATAGACGAACTTCTTGTATATGACATACCAACTTCGATGACGCTTATCTGGGA 449
Qy 439 TCTGCTCAGCGCTTCAAGATGCTATGAGAGCTGCTCAAGCTGGAAGAGTCCAGCT 498
Dh 450 CCAACCAATCTTTGGAGACGAAATATGTGTGCTTTCTAAGCCCTTAAATCGTGA- 506
Qy 499 GGAATTGACGCGATATGCGAGGTGTGCTTGAACGTTTGGGACCGATCCGACATG 558
Dh 507 AGATTTGAGATTTTCAAGAAATTTGTGTAAGACTTTGGAGATTAAGTTAAGATGTG 566
Qy 559 GATCACCATCAACGACCCCTGATTCAGGCATCTATGATATGCCACCGGACGCAACGC 618
Dh 567 GACAAACATCAACGACCTTATATATGACGTGTGGGTTATGATCAAGGTAAACAAGGC 626
Qy 619 CCGGCGAAGAGC-----AGCATTAACAACACTCCACGAGGGCAACACTGCCACTGA 672
Dh 627 GCGTGAACGATGTCTCAAAATGGGTAAACGAAAGGTGCAAGCTGGAGATTCGATACGA 686
Qy 673 GCGGTGCTGCTGGGAAAGGCCAGATCATGAGCATGCCGCGCGTGGCCCTTCAAG 732
Dh 687 GCTTACATTTGTTCACATCACTCTTCTTCCCATGCCGCTGCAATGAAAGATTTTCG 746
Qy 733 CAGGACCTTTGCCCCCTGCAAAAGGCGCAAGATCGGATCTCGTCAACGCGCATCTA 792
Dh 747 AAATGTGAAAAAATCTTCGATGATGCGCAATTTGGATGATATCAACCAAGATGTT 806
Qy 793 TGAACCTTGGACACGATGACCTCGGACACAGAGGCTGTAGGACCGATGGAATT 852
Dh 807 CGACCTTATATCTTCGATTCACATGACGATTAAGAGACGCTAACGAGCTTTCCTT 866
Qy 853 TCACATTTGGCTGTTTGGCAATCCCATCTTCTTGAAGAGGACTATCCAGAGACATGAA 912
Dh 867 TGAATTTGATGATGATCTTGTATCCAGTATTC-----ACGAGATTTATCCAGATTTGTA 923
Qy 913 GACAGAGCTGGGACGAGGCTTCCAGCCCTCATCTCCGCGACTTGGCATCTCAATGC 972
Dh 924 AAGATGACGCGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 983
Qy 973 CGGAGACGCACTTTACGCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1032
Dh 984 TTCA---TCAGATTTTGTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1040
Qy 1033 CGGTCCGCTCCCGAGACGACTATCTCGGCGCCATCATGA-----GCACACGA 1083
Dh 1041 TCACTATGACCTCTGAAAAAATCTGTTTCAAACTGACCAACATGTGAATGAAATGTA 1100
Qy 1084 GAATTAAGACGCGACGCCCTGTGGCGAGAGAGCGGCTGCTGCGCTGCTGCTGCTG 1143
Dh 1101 TAATCAAGTGTGATCATCATCGACCTGCGGAAGAAAGGCGCTTCTTAATTTTCAAC 1160
Qy 1144 GCAATGTTCCGGAAGCATCTCGCCCGGCTGATAGGCGCTGAC---GGCAAGCCATCT 1200
Dh 1161 GGAAGGCTTACGAAAGTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1220
Qy 1201 CATCAACGAGAGGATGCCGCTGAGAGAGAGAACTGACGTGCGAGAGGCGCT 1260
Dh 1221 CATCAAGAAATATGATTAATTAATGACACGACGATCAAAACAGAGAGAAATTTGT 1280
Qy 1261 CACGACCCCTTCCGATCCGCTATCTTGAATCTGACATCTGATCTGATTTTCAAGGCAT 1320
Dh 1281 GAGGACACATTTTGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1340
Qy 1321 TACCAAGACGCGCTGCTGCAAGGGTACTTTGCGTGGCGCTTGCATTAATCTTGA 1380

Db 1359 TGGTGGCAAAAGTATTGGATATCTTGTGGTCAATGCTTGACACTTTCAGATGGAGGCT 1418
Oy 1390 TGGCTACGAGACCCAGATTCCGCGCTCAGCTTCACAGACTACACCCCTCAAGCCGACGCC 1449
Db 1419 CGGGTACACTTCCCGCTTTGGCATCGTCTAGCTGAGACGCTAAAGAGGTACCC 1478
Oy 1450 CAAGAACTGCGCCCTGGTCTCTCAAGACATGTTGGCGCCCGGACAGG 1498
Db 1479 CAAGACTACACTTCTGTGTTCAAGAACATGCTCTCCAGTAAGAGAGG 1527
RESULT 11
AAS21370
ID AAS21370 standard; cDNA; 2016 BP.
XX AAS21370;
XX
XX
XX 24-OCT-2001 (first entry)
XX
XX
DE Human cDNA sequence encoding for PRO9820 polypeptide.
XX
XX Human secretory and transmembrane; PRO; mammalian; cancer; lung; breast;
KW prostate; cervical; tumour necrosis factor-alpha; TNF-alpha; cartilage;
KW ear; proliferation; glucose; free fatty acid; skeletal muscle; adipocyte;
KW A-peptide; factor VIIA; gene therapy; bs.
XX
XX Homo sapiens.
XX
XX MO200140466-A2.
XX
XX 07-JUN-2001.
XX
XX 01-DEC-2000; 2000WO-US032678.
XX
XX 01-DEC-1999; 99WO-US028301.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.
PR 02-DEC-1999; 99WO-US028564.
PR 02-DEC-1999; 99WO-US028565.
PR 09-DEC-1999; 99US-0170262P.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 20-DEC-1999; 99WO-US030999.
PR 30-DEC-1999; 99WO-US031274.
PR 30-DEC-1999; 99WO-US031275.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000277.
PR 06-JAN-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 18-FEB-2000; 2000WO-US004342.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US005004.
PR 01-MAR-2000; 2000WO-US005601.
PR 02-MAR-2000; 2000WO-US005841.
PR 03-MAR-2000; 2000US-0187202P.
PR 10-MAR-2000; 2000WO-US006319.
PR 15-MAR-2000; 2000WO-US006884.
PR 20-MAR-2000; 2000WO-US007377.
PR 21-MAR-2000; 2000WO-US007532.
PR 30-MAR-2000; 2000WO-US008439.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 05-JUN-2000; 2000US-0209832P.
PR 28-JUL-2000; 2000WO-US020710.
PR 11-AUG-2000; 2000WO-US022031.
PR 23-AUG-2000; 2000WO-US023522.
PR 24-AUG-2000; 2000WO-US023328.
PR 08-NOV-2000; 2000WO-US030952.

PR 10-NOV-2000; 2000WO-US030873.
XX
XX (BETH) GENENTECH INC.
XX Baker KP, Beresini M, Deforge L, Desnoyers L, Flivaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski P, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
XX MPI: 2001-408281/43.
DR P-PSDB; AAU12238.
XX
XX Isolated , secretory and transmembrane PRO polypeptide used to detect
PT other PRO polypeptides, link bioactive molecules to cells expressing PRO
PT polypeptides, and detect the presence of mammalian tumors e.g. lung,
PT breast, prostate, cervical.
XX
XX Claim 3, Fig 253; 813pp; English.
XX
XX AAS21244-AAS21518 encode for novel human secretory and transmembrane PRO
XX polypeptides. The PRO polypeptides are useful to detect other PRO
CC polypeptides, to link bioactive molecules to cells expressing PRO
CC polypeptides, to modulate biological activities of cells expressing PRO
CC polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample. Some
CC of the 275 sequences are also useful to stimulate the release of tumour
CC necrosis factor-alpha (TNF-alpha) from human blood, the proliferation or
CC differentiation of chondrocytes, the proliferation or gene expression in
CC pericyte cells, the release of proteoglycans from cartilage, the
CC proliferation of inner ear utricular supporting cells or of T-
CC lymphocytes, the release of a cytokine from peripheral blood monocytes
CC (PBMCs), or the proliferation of endothelial cells. Some of the PRO
CC polypeptides may modulate glucose or free fatty acid uptake by skeletal
CC muscle cells or by adipocytes; or inhibit binding of A-peptide to factor
CC VIIA. The PRO polypeptides can be used in assays to identify molecules
CC involved in binding interactions. The polynucleotides encoding PRO
CC polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy
XX
XX Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;
SQ
Query Match 9.0%; Score 178.2; DB 4; Length 2016;
Best Local Similarity 54.5%; Pred. No. 1.2e-33;
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;
Oy 76 AGCTCTGCCCAACGACTTTGATGAGGCTTGCACACGCGCTTACCGATGAGGCGC 135
Db 237 AACCTTCCCTTTGGCTTCTTCCTGGGCGTGGGCAAGTTCTGCTTACAGAGGAGGCGC 296
Oy 136 CGTCAAGAGAGTGCGCGCGCGCGCTTCATTTGGGAACAGTACCTGAGGCGATC 195
Db 297 CTGGGACACGAGACGGGAAGGGCTTACGATCTGGGAAGCTTTCACACAGTGGGAAGG 356
Oy 196 GCGCACCAAGG---CGCAACGGCGATGAGGCTTGGGATCTACCAACGCTACGATGA 252
Db 357 GAAAGTGTGGGAATGAGACGAGAGTATGCTGTGACGGCTACTACCAAGGTCACGA 416
Oy 253 GAACTTGATCTTTGACCAAGTACGCGGCAAGGCTTACCGCTTCTTGTGTGTC 312
Db 417 GGAATCATTTGTGCTGAGGGAACGTGACCACTACCGATTCCTCGTGTGGCC 476
Oy 313 GCGGATCTTCCCTCGCGGCGAGCTGAGTCCCGTCAACGAGAGGGAATTGAGTTTA 372
Db 477 CCGGCTCTGCGCCACAGGCAATCGAGCGGACAGGTGAACGAAGGGAATGAAATCTTA 536
Oy 373 CAGCAAACTGATTGACGCGCTGTTGAGCGGGGATATCACGCTTGGGATCTTTGTAACA 432
Db 537 CAGTATCTTATTCATGATGCGCTTCTGAGACGAACATCACTCCATCGTGAACCTTGACA 596
Oy 433 CTGGGATCTGCTCAGGCGCTTACAGATCGATTATGAGAGCTGTGCTCAACGTGGAAGAGT 492
Db 597 CTGGGATCTGCGACAGCTGTCTCAAGTCAATACGATGAGTGGTGAAGATGTGAGCATGGC 656

QY 493 CCACTGACCTTTAGAGCGGATATGCGAGTTGTCTTTGAACGTTTGGGAGCCAGTCA 552
 Db 657 CAA---CTACTTCAAGAGACTAAGCCAACTGTGCTTTGAGCCCTTTGGGACCGTGTGA 713
 QY 553 GAACTGATACCATCAACGACCCCTTGATTCAGGCCATTTATGATATATCCACCGGAC 612
 Db 714 GCATCGATATCACTTATGATATCTCGGGCAATGGCAGAAAAGGCTATGAGACGGGCCA 773
 QY 613 CAACGCCCCGGGAGGAGCAGCATTTAACAAGCACTCCACGAGGGCAACCTGCACTGA 672
 Db 774 CATGGCGCGGAGCTTAAGCTCCG-----GGCACGG 806
 QY 673 GCGGTGCTGCTGTAAGAGCCGAGATCATAGCCATGCGCGCGCTGCGCTTACAG 722
 Db 807 CCGTCAAGAGGACACACACATCATTTAAGGCCACGCAAAACCTGGCATTTCTTATA 866
 QY 733 CAGGACCTTTCGCCCCCTCGCAAAAGGCCAGATGCGCATTCGCTCAACGGCACTACTA 792
 Db 867 CACCACTGGCGGAGCAAGAGTCTGTGGGAATTTCACTGAATGTGACTGGGG 926
 QY 793 TGAAGCCCTGGGAGAGCAATGAGCTGGGAGCAAGAGGCTGAGAGGAGATGGAAT 852
 Db 927 GGAACCTGTGACATTTAGTAACTTAAAGGACCTAGAGGCTGCGAGAGATACCTACGTT 966
 QY 853 TCACATTGGCTGTTTGCATCCATCT 881
 Db 987 CTGCTGGGCTGTTTGCATCCATTT 1015

RESULT 12
 ABL88245
 ID ABL88245 standard; cDNA; 2016 BP.

XX ABL88245;

XX 16-MAY-2002 (first entry)

DE Human PRO9820 cDNA sequence SEQ ID NO:147.

XX Human; angiogenesis; cardiant; cytosolic; antiangiogenic; hypotensive;
 KW vulnary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
 KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
 KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
 KW age-related macular degeneration; arterial restenosis; angina;
 KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
 KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
 KW wound healing; chromosome mapping; gene mapping; gene; 88.

OS Homo sapiens.

XX MO200200690-A2.

PD 03-JAN-2002.

PF 20-JUN-2001; 2001WO-US019692.

XX 23-JUN-2000; 2000US-0213637P.
 PR 20-JUL-2000; 2000US-0219556P.
 PR 25-JUL-2000; 2000US-0220624P.
 PR 28-JUL-2000; 2000US-0220641P.
 PR 02-AUG-2000; 2000US-0222655P.
 PR 17-AUG-2000; 2000US-00643657.
 PR 23-AUG-2000; 2000WO-US023522.
 PR 24-AUG-2000; 2000WO-US023328.
 PR 07-SEP-2000; 2000US-023078P.
 PR 18-SEP-2000; 2000US-00664610.
 PR 24-OCT-2000; 2000US-0242922P.
 PR 08-NOV-2000; 2000US-0070923P.
 PR 08-NOV-2000; 2000WO-US030873.
 PR 10-NOV-2000; 2000WO-US030873.
 PR 01-DEC-2000; 2000WO-US032678.

PR 20-DEC-2000; 2000US-00747259.
 PR 20-DEC-2000; 2000WO-US034956.
 PR 22-JAN-2001; 2001US-00767609.
 PR 28-FEB-2001; 2001US-00796498.
 PR 28-FEB-2001; 2001WO-US005520.
 PR 01-MAR-2001; 2001WO-US006666.
 PR 09-MAR-2001; 2001US-00802706.
 PR 14-MAR-2001; 2001US-00806889.
 PR 22-MAR-2001; 2001US-00815744.
 PR 05-APR-2001; 2001US-00828366.
 PR 10-MAY-2001; 2001US-00854208.
 PR 10-MAY-2001; 2001US-00854280.
 PR 25-MAY-2001; 2001US-0086028-
 PR 25-MAY-2001; 2001US-0086034-
 PR 25-MAY-2001; 2001WO-US017092.
 PR 30-MAY-2001; 2001US-00870574.
 PR 30-MAY-2001; 2001WO-US017443.
 PR 01-JUN-2001; 2001WO-US017800.

(GETH) GENENTECH INC.

PI Baker KP, Ferrara N, Gerber H, Gertlisen ME, Goddard A,
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;

DR WPI; 2002-090516/12.
 DR P-PSDB; ABB84990.

PT One hundred and eighty seven nucleic acid encoding PRO polypeptides,
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal.

XX Claim 2; Fig 347; 565pp; English.

PS ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
 CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytosolic,
 CC antiangiogenic, hypotensive, vulnary and antiarteriosclerotic
 CC activities, and can be used in gene therapy. The PRO polynucleotides,
 CC proteins, agonists and antagonists are useful for treating or diagnosing
 CC a cardiovascular, endothelial or angiogenic disorder in a mammal, e.g.
 CC cardiac hypertrophy, trauma, cancer, age-related macular degeneration,
 CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
 CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
 CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
 CC healing. The PRO polynucleotides have applications in molecular biology,
 CC including use as hybridisation probes, and in chromosome and gene
 CC mapping. ABL88259 to ABL88267 represent primers and probes used in the
 CC exemplification of the present invention

XX Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

XX Query Match 9.0%; Score 178.2; DB 6; Length 2016;

XX Best Local Similarity 54.5%; Pred. No. 1.2e-33;

XX Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

QY 76 AGCTGCCCAAGCATTTGAATGGGCTTCCGAAGCGCGCTTACAGATGAGGCGC 135
 Db 237 AATCTTCCCTCTTGGCTTCTCTGGGCGTGGGCACTTCTGCTTACAGAGGCGC 236
 QY 136 CGTCAAGAGGTCGCGCGCGCCCTTCATCTGGGACAGTACTGCGACCTGAGCCATC 195
 Db 297 CTGGGACGAGGAGGAGAAAGGCGCTAGCATCTGGGACGCTTACACACAGTGGAGG 356
 QY 196 GCGGACCAACG--CGCAACGCGCATGTGGCTTGCATCTACACCGCATCATGA 252
 Db 357 GAAAGTCTGGGAATGAGCGGCAATGATGAGCTGTGACGCTACTACAAAGTCCAGA 416
 QY 253 GACATTTGATCTCTTGACCAAGTACGCGCAAGGCGCTTCCCTGCTGGTGTGTC 312
 Db 417 GGCATCATTTCTCTAGGAGGAGTGCAGTCAACCTACCAATTTCTCTGCTTGGCC 476
 QY 313 GCGGATCATTTCCCTCGCGGCGAGGCTGATCCCGTCAACGAGAGGAAATGATTTTA 372

```

Db 477 CGGCTCTGCCCCACAGGATCCGAGCCGAGAGGTGAACAAGAGGATTCATTTCTA 536
Qy 373 CAGCAATGATGATGACGCCCTTGTGAGGGGGGATTCACGCCCTTGGGATCTTGTACCA 432
Db 537 CAGTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 596
Qy 433 CTGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 492
Db 597 CTGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 656
Qy 493 CCAGCTGATCTTGAAGCGGTATGAGAGGTGTGCTTTGAACGTTTGGGAGCCGAGTCCA 552
Db 657 CAA---CTACTTCAGAGACTAGCCCACTGCTGCTTGTAGGCTTTGGGAGCCGCTGTA 713
Qy 553 GAACCTGATTCACATCAAGAACCCCTGATTCAGGCCATTCATGATATGCCACCGGAG 612
Db 714 GATCTGATTCACGTTTCACTGATCTCGGGCAATGCGCAAAAAGGCTATGAGACGGGCA 773
Qy 613 CAACGCCCCGGGCAAGAGCAGCATTTAACAGCACTCCACGAGGGCAACATGCCACTGA 672
Db 774 CCATGCGCGGGGCTCGAAGCTCCGC-----GGCACCGG 806
Qy 673 GCCCTGCTGCTGGAAGGCCAGATATAGCCATGCCCGCGCTGCTGCTACAG 732
Db 807 CCTGTACAAAGGACACACATCATTTAAGGCCACCAAACTGGCACTTCTTATTA 866
Qy 733 CAGGATCTTTCGCCCCCTCGCAAAAGGGCCAGATGGGATCTCGCTCAACGGCGACTCA 792
Db 867 CACCACTGCTGCTGCAAGCAAGCAAGCTTCTGTGGAAATTCATGAACTGTGACTGGG 926
Qy 793 TGAAGCTTGGGAGCAATGAGCTCGGAGCAAGAGGCTGCTGAGCGACGATGGAATT 852
Db 927 GGAACCTGTGAGCACTAGTATGATCCCAAGAGCCTAGAGCTGCCAGAGATACCTACGTT 986
Qy 853 TCACATGGCGCTGTTGGCAATCCCATCT 881
Db 987 CTGTCTGGCTGTGTTGCCAACCATT 1015

RESULT 13
ABL95734
ID ABL95734 standard; cDNA; 2016 BP.
XX
AC ABL95734;
XX
DT 19-JUL-2002 (first entry)
XX
DE Human angiogenesis related cDNA PRO9820 SEQ ID NO: 347.
XX
KW Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiatic; cytosolic; antiangiogenic; hypotensive; vulnerary;
KW antiarteriosclerotic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200208284-A2.
XX
PD 31-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US021735.
XX
XX 20-JUL-2000; 2000US-0219556P.
XX 25-JUL-2000; 2000US-0220664P.
XX 28-JUL-2000; 2000WO-US020710.
XX 02-AUG-2000; 2000US-0222695P.
XX 17-AUG-2000; 2000US-00643657.
XX 23-AUG-2000; 2000WO-US023328.
XX 24-AUG-2000; 2000WO-US023322.
XX 07-SEP-2000; 2000US-0230978P.
XX 18-SEP-2000; 2000US-00664610.
XX 18-SEP-2000; 2000US-00665350.

```

```

PR 24-OCT-2000; 2000US-0242922P.
PR 08-NOV-2000; 2000US-00709238.
PR 08-NOV-2000; 2000WO-US030952.
PR 10-NOV-2000; 2000WO-US030873.
PR 01-DEC-2000; 2000WO-US032678.
PR 20-DEC-2000; 2000US-00747259.
PR 20-DEC-2000; 2000WO-US034956.
PR 22-JAN-2001; 2001US-00767609.
PR 28-FEB-2001; 2001US-00796498.
PR 28-FEB-2001; 2001WO-US006520.
PR 01-MAR-2001; 2001WO-US006666.
PR 09-MAR-2001; 2001US-00802706.
PR 14-MAR-2001; 2001US-00806889.
PR 22-MAR-2001; 2001US-00816744.
PR 05-APR-2001; 2001US-00828366.
PR 10-MAY-2001; 2001US-00854208.
PR 10-MAY-2001; 2001US-00854280.
PR 25-MAY-2001; 2001US-00866028.
PR 25-MAY-2001; 2001US-00866034.
PR 25-MAY-2001; 2001WO-US017092.
PR 30-MAY-2001; 2001US-00870574.
PR 30-MAY-2001; 2001WO-US017443.
PR 01-JUN-2001; 2001WO-US017800.
PR 20-JUN-2001; 2001WO-US019692.
XX

```

```

PA (GETH ) GENENTECH INC.
PA (BAKE/) BAKER K P.
PA (FERR/) FERRARA N.
PA (GERB/) GERBER H.
PA (GERR/) GERRIESEN M E.
PA (GODD/) GODDARD A.
PA (GODO/) GODOWSKI P J.
PA (GURN/) GURNEY A L.
PA (HILL/) HILLMAN K J.
PA (MARS/) MARSTERS S A.
PA (PANJ/) PAN J.
PA (PAON/) PAONI N F.
PA (STEP/) STEPHAN J F.
PA (WATA/) WATANABE C K.
PA (WILL/) WILLIAMS P M.
PA (WOOD/) WOOD W I.
XX

```

```

PI Baker KP, Ferrara N, Gerber H, Gerriksen ME, Goddard A,
PI Godowski PJ, Gurney AL, Hillman KJ, Marsters SA, Pan J, Paoni NF,
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W,
XX WPI, 2002-171999/22.
DR P-PSDB; ABB95596.
XX

```

```

PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal.
XX

```

```

PS Claim 1; Fig 347; 567bp; English.
XX

```

```

XX The present invention provides the protein and coding sequences of human
XX PRO proteins. These are useful for treating or diagnosing a
XX cardiovascular, endothelial or angiogenic disorder, including cardiac
XX hypertrophy, trauma, cancer, age-related macular degeneration,
XX atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
XX angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
XX angiogenesis (such as breast carcinoma and liver carcinoma) and wound
XX healing. The present sequence is a coding sequence of the invention
XX

```

```

SQ Sequence 2016 BP, 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

```

```

Query Match 9.0%; Score 178.2; DB 6; Length 2016;
Best Local Similarity 54.5%; Pred. No. 1.2e-33;
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

```

```

Qy 76 ACCTCTGCCCAAGCATTTGAATGGGCTTGGCAAGCGCCGCTTACCAAGTGAAGAGCGC 135
Db 237 AACCTTCCCTTGTGCTTCTCTGCGGGCGTGGGCGAGTTCTGCTTACCAAGAGAGCGCG 296

```

QY 136 CGTCAAGAAGTGGCCGCGCCCGCTCATCTGGGACACGACTGCCACCTGAGCCATC 195
DB 297 CTGGGACCAAGACCGGAAAGGGCTTACATCTGGGAGCTTTCACACACACTGGGAAGG 356
QY 196 GCGGACCAACGG---CGCCAAACGGCGATGTGGCTTGGCATCTACCAACCGTACGATGA 252
DB 357 GAAAGTCTTGGGAATGAGACGGGAGATGTAGCTGTGAGCGGTACTACAAAGGTCCAGGA 416
QY 253 GGACTTGTATCTCTTGACCAAGTACGGCGCAAGGCGCTTACCGCTTCTCTTGTGTGTC 312
DB 417 GGACATCATCTCTCTGGGGAACGTACGCTACCACTACCGATTCCTCCCTGCTTGGCC 476
QY 313 GCGGATCATCTCCCTCGCGGCGGACAGCTGATCCCGTCAAGAGAGGGAATTGATTTTA 372
DB 477 CCGGCTCTGCGCCACAGGCACTCCAGCGGACAGGTAAACAAGAGGAATCAATTCTA 536
QY 373 CAGCAAACTGATTGACCGCCCTGTTGAGCGGGATATCACCGCTTGGGTGACTTTTACA 432
DB 537 CAGTGATCTTATGATGCTCCCTTCTGACAGCAATCATCTCCCATCTTGACCTTGACCA 596
QY 433 CTGGGATCTGCTTACGGCGCTTACGATCGCTATGAGAGCTGGCTCAACGTGGAAAGGT 492
DB 597 CTGGGATCTGCAACGCTGCTCCAGGTCAATACGCTGGGTGGAGATGTAGCATGCG 656
QY 493 CCAAGCTGACTTTGAGGGGATGCGAGGTTGTGCTTTGAAAGTTTGGGGACGAGTCCA 552
DB 657 CAA---CTACTTCAAGACCTACGCGCAACCTGTGCTTTGAGGCTTTGGGAGCCTGTGAA 713
QY 553 GAATGATGATCAGCATCAACGACGACGCTGATTCAGGCGCATCTATGATATGCAACCGGAC 612
DB 714 GCATGATCAGCATCGTTCAGTGTCTCTCGGGCAATGGCAAAAGGCTATGAGAGGGCCA 773
QY 613 CAACGCCCCGGGCGAGGACGATTAACAGCACTCCACGAGGGCAACACTGCACTGA 672
DB 774 CCAATGCCCGGGCGCTGAAGCTCCG-----GGCACCGG 806
QY 673 GCGGCTGCTGCTGGAAGGCGCCAGATCATGAGCCATGCCGCGCGCTGCGCTACAG 732
DB 807 CTTGTACAGAGCACACACACATCATTTAAGCCACGCAAACTTGCACTTTTATTA 866
QY 733 CAGGACCTTTGCGCCCTTCGCAAAAGGCGCAGATCGGATCTGCTCAACGCGACTACTA 792
DB 867 CACCACTGGCGGACGAAAGGACGAAAGTCTGTGGAAATTTCACTGAATGTGTGCTGGG 926
QY 793 TGAGCCCTGGGACAGCAATGACCTCGGACAGAGGCTGTGAGCGAGATGAATT 852
DB 927 GGAACCTGTGACATTAATGAAACCCCAAGGACCTAGAGGCTGCCAGAGATACCTACATT 986
QY 853 TCACATTGGCTGTTGCCAATCCCATCT 881
DB 987 CTGCTGGGCTGGTTGCCAACCCTATT 1015

RESULT 14
ACD23979 standard; cDNA; 2016 BP.
ACD23979;
DT 26-AUG-2003 (first entry)
XX Novel human secreted and transmembrane protein PRO9820 cDNA.
XX Human; secreted and transmembrane protein; PRO; anti-inflammatory;
KM antiarteriosclerotic; cardiant; anti-infertility; anti-HIV; cytostatic;
KM anti-diabetic; gene therapy; tumour necrosis factor (TNF)-alpha release;
KM TNF-alpha release; cell proliferation; cell differentiation;
KM gene expression modulator; proteoglycan release; cytokine release;
KM tumour; inflammatory disease; organ failure; atherosclerosis;
KM cardiac injury; infertility; birth defect; premature aging; AIDS;
KM acquired immunodeficiency syndrome; cancer; diabetic complication;
KM chromosome mapping; gene mapping; pharmaceutical; diagnostic; biosensor;

KM bioreactor; tissue typing; gene; ss.
XX Homo sapiens.
OS US2003032156-A1.
XX 13-FEB-2003.
PD 06-MAY-2002; 2002US-00140474.
PF
XX 31-MAR-1997; 97WO-US005230.
XX 12-JUN-1998; 98WO-US012456.
PR 14-JUL-1998; 98WO-US014552.
PR 28-AUG-1998; 98WO-US017888.
PR 10-SEP-1998; 98WO-US018824.
PR 14-SEP-1998; 98WO-US019093.
PR 14-SEP-1998; 98WO-US019094.
PR 14-SEP-1998; 98WO-US019177.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98WO-US019437.
PR 07-OCT-1998; 98WO-US021141.
PR 29-OCT-1998; 98WO-US022991.
PR 29-OCT-1998; 98WO-US022992.
PR 20-NOV-1998; 98WO-US024855.
PR 01-DEC-1998; 98WO-US025108.
PR 05-JAN-1999; 99WO-US000106.
PR 08-MAR-1999; 99WO-US005028.
PR 10-MAR-1999; 99WO-US005190.
PR 20-APR-1999; 99WO-US008615.
PR 14-MAY-1999; 99WO-US010733.
PR 02-JUN-1999; 99WO-US012252.
PR 01-SEP-1999; 99WO-US020111.
PR 08-SEP-1999; 99WO-US020594.
PR 13-SEP-1999; 99WO-US020944.
PR 15-SEP-1999; 99WO-US021090.
PR 15-SEP-1999; 99WO-US021547.
PR 05-OCT-1999; 99WO-US023089.
PR 29-NOV-1999; 99WO-US028214.
PR 30-NOV-1999; 99WO-US028313.
PR 30-NOV-1999; 99WO-US028409.
PR 01-DEC-1999; 99WO-US028301.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.
PR 02-DEC-1999; 99WO-US028564.
PR 02-DEC-1999; 99WO-US028565.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 20-DEC-1999; 99WO-US030999.
PR 22-DEC-1999; 99WO-US030720.
PR 30-DEC-1999; 99WO-US031243.
PR 30-DEC-1999; 99WO-US031274.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000277.
PR 11-FEB-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 18-FEB-2000; 2000WO-US004342.
PR 22-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US004914.
PR 01-MAR-2000; 2000WO-US005004.
PR 02-MAR-2000; 2000WO-US005046.
PR 10-MAR-2000; 2000WO-US005841.
PR 10-MAR-2000; 2000WO-US006319.
PR 15-MAR-2000; 2000WO-US006884.
PR 20-MAR-2000; 2000WO-US007377.
PR 21-MAR-2000; 2000WO-US007532.
PR 30-MAR-2000; 2000WO-US008439.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 28-JUL-2000; 2000WO-US020710.

PR 11-AUG-2000; 2000WO-US022031.
 PR 23-AUG-2000; 2000WO-US023522.
 PR 24-AUG-2000; 2000WO-US023328.
 PR 08-NOV-2000; 2000WO-US030952.
 PR 10-NOV-2000; 2000WO-US030873.
 PR 01-DEC-2000; 2000WO-US032678.
 PR 20-DEC-2000; 2000US-00747259.
 PR 20-DEC-2000; 2000WO-US034956.
 PR 28-FEB-2001; 2001US-00796498.
 PR 28-FEB-2001; 2001WO-US006520.
 PR 01-MAR-2001; 2001WO-US006666.
 PR 09-MAR-2001; 2001US-00802706.
 PR 14-MAR-2001; 2001US-00808689.
 PR 22-MAR-2001; 2001US-00816744.
 PR 05-APR-2001; 2001US-00828366.
 PR 10-MAY-2001; 2001US-00854208.
 PR 10-MAY-2001; 2001US-00854280.
 PR 18-MAY-2001; 2001US-00860216.
 PR 25-MAY-2001; 2001US-00866028.
 PR 25-MAY-2001; 2001US-00866034.
 PR 25-MAY-2001; 2001WO-US017092.
 PR 01-JUN-2001; 2001US-00872035.
 PR 05-JUN-2001; 2001US-00871800.
 PR 14-JUN-2001; 2001US-00874503.
 PR 19-JUN-2001; 2001US-00886342.
 PR 20-JUN-2001; 2001WO-US019692.
 PR 21-JUN-2001; 2001US-00887879.
 PR 22-JUN-2001; 2001WO-US020116.
 PR 29-JUN-2001; 2001WO-US021066.
 PR 09-JUL-2001; 2001WO-US021735.
 PR 18-JUL-2001; 2001US-00908827.
 PR 06-AUG-2001; 2001US-00924419.
 PR 09-AUG-2001; 2001US-00927796.
 PR 16-AUG-2001; 2001US-00931836.
 PR 19-DEC-2001; 2001US-00028072.
 (GENTH) GENENTECH INC.
 XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerlicsen ME, Goddard A, Godowski RJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tunas D, Watanabe CK, Wood WI, Zheng Z;
 XX WPI: 2003-341980/32.
 DR P-PSDBJ; ABO17742.
 XX
 PT New secreted and transmembrane PRO nucleic acids, for treating
 PT inflammation, organ failure, atherosclerosis, cardiac injury,
 PT infertility, birth defects, premature aging, acquired immunodeficiency
 PT syndrome (AIDS), or cancer.
 XX
 XX Claim 2; Fig 253; 660pp; English.
 PS
 XX The invention describes an isolated nucleic acid (I) comprising, or which
 CC has 80 % sequence identity to, or the full-length coding sequence of, one
 CC of 275 nucleotide sequences, and which encodes a corresponding
 CC polypeptide selected from 275 amino acid sequences, where all sequences
 CC are given in the specification. The polypeptide encoded by (I) is used to
 CC detect PRO polypeptides, link a bioactive molecule to a cell expressing a
 CC PRO polypeptide, modulate a biological activity of a cell, stimulate the
 CC release of tumour necrosis factor (TNF)-alpha from human blood, modulate
 CC the uptake of glucose or free fatty acid by cells, stimulate or inhibit
 CC the proliferation or differentiation of cells or gene expression,
 CC stimulate the release of proteoglycans, stimulate the release of cytokine
 CC from peripheral blood mononuclear cells, inhibit the binding of A-peptide
 CC to factor VIIa, or detect the presence of tumour in a mammal. The nucleic
 CC acid and polypeptide encoded by it, are useful for treating inflammatory
 CC diseases, organ failure, atherosclerosis, cardiac injury, infertility,
 CC birth defects, premature aging, acquired immunodeficiency syndrome
 CC (AIDS), cancer, or diabetic complications. The nucleic acid is useful as
 CC hybridisation probes, in chromosome and gene mapping, and in generating
 CC antisense RNA or DNA. The polypeptides are useful as pharmaceuticals,
 CC diagnostics, biosensors or bioreactors. Both are useful in tissue typing.

CC This sequence encodes a novel human secreted and transmembrane PRO
 CC polypeptide
 XX
 SQ Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;
 Query Match 9.0%; Score 178.2; DB 7; Length 2016;
 Best Local Similarity 54.5%; Pred. No. 1.2e-33;
 Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;
 QY 76 AGCTCTGCCCAACACCTTTGAATGGGGCTTCGCAACGCCGCCCTACCGATTCGAGCGC 135
 DB 237 AACCTTCCCTTGGCTTCTCTGGGGGCTGGGAGATTCTGCTTACGACGAGGGCGC 236
 QY 136 CGTAAAGAGGCG 135
 DB 297 CTGGGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 356
 QY 196 GCGCACCAACG--CGSCAACGCGCATGTGGCTTGGATCACTAACACCGCTACGATGA 252
 DB 357 GAAAGTCTTGGGAATGAGACGGCAGATGAGCTGTGAGCGCTTACTACAGGTCCAGGA 416
 QY 253 GGAATTGATCTCTTGACCAAGTACGCGCAAGGCGCTTACCGCTTCTCTGTGTGTC 312
 DB 417 GGAATCATTTCTGCTGAGGAACTGCACTGCACTACCGATTCCTCTGTGTGCGC 476
 QY 313 GCGGATCATTCCTCG 372
 DB 477 CGGCTCTGCGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 536
 QY 373 CAGCAACTGATTTAGACCGCTCTTGAAGGCGGAGGATCACGCTTGGGTGACTTGTACCA 432
 DB 537 CAGGATCTTATGAGAGCCCTTGTGAGCAGCAATCACTCCCATGTGACTTGGACCA 596
 QY 433 CTGGGATCTGCTCAGCGCTTACAGATGCTATGAGAGGTGCTCAAGTGAAGAGGT 492
 DB 597 CTGGGATCTGCGCACACTGCTCCAGGTCAAAATCGGTGGGTGACAGATGTGAGCATGCG 656
 QY 493 CCACTGGAATTTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552
 DB 657 CAA---CTACTTACAGACTACGCAACTGTGCTTGAAGCCTTGGGAGCCGTGTGA 713
 QY 553 GAACTGATCAGCATCAAGAGCCCTGATTCAGGCGCATCTATGATATGATCCACCGCAG 612
 DB 714 GCACTGATCAGCTTCAATGATCTCGGGCAATGCGAAGGAGGAGGAGGAGGAGGAGG 773
 QY 613 CAACGCCCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 672
 DB 774 CAAATGCGCGGCGCTGAAGCTCGC-----GGCACCGG 806
 QY 673 GCGGTGCTCGCTGAGAAAGCCAGATCATGAGCCATGCGCGCGCGCGCGCGCTTACAG 732
 DB 807 CCGTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 866
 QY 733 CAGGACTTTCGCGCTCGCAAAAGGCGCAATCGGATCTGCTCAACGCGGACTACTA 792
 DB 867 CACCAAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 926
 QY 793 TGAAGCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 852
 DB 927 GGAACCTTGGAGATTTAGTAAACCCCAAGGAGCTTGAAGGCTGCGAGATTAAGTCA 986
 QY 853 TCACATTTGCTGTTTTCATTCCTCATCT 881
 DB 987 CTGTCTGGGCTGTTTTCGCAACCCCATTT 1015
 RESULT 15
 ACA67120
 ID ACA67120 standard; cDNA; 2016 BP.
 XX ACA67120;
 AC
 XX 23-JUN-2003 (first entry)

XX cDNA encoding human PRO polypeptide #127.
DE
XX
KM Human; PRO polypeptide; secreted and transmembrane protein;
KM anti-PRO antibody; diagnostic assay; gene expression; diabetes;
KM bone disorder; cartilage disorder; rheumatoid arthritis; obesity;
KM sports injury; osteoarthritis; hyper-insulinaemia; hypo-insulinaemia;
KM hearing loss; coagulation disorder; stroke; heart attack; cardiac;
KM antidiabetic; anorectic; vulnery; antiarthritic; osteopathic;
KM antirheumatic; auditory; cerebroprotective; angiogenic; gene; ss.
OS
XX Homo sapiens.
PN US2003004311-A1.
XX
PD 02-JAN-2003.
XX
PF 19-DEC-2001; 2001US-00028072.
XX
XX 18-JUN-1997; 97US-0049911P.
PR 26-AUG-1997; 97US-0056974P.
PR 17-SEP-1997; 97US-0059113P.
PR 17-SEP-1997; 97US-0059115P.
PR 17-SEP-1997; 97US-0059117P.
PR 17-SEP-1997; 97US-0059122P.
PR 17-SEP-1997; 97US-0059184P.
PR 18-SEP-1997; 97US-0059263P.
PR 19-SEP-1997; 97US-0059352P.
PR 19-SEP-1997; 97US-0059588P.
PR 24-SEP-1997; 97US-0059836P.
PR 17-OCT-1997; 97US-0062250P.
PR 17-OCT-1997; 97US-0062285P.
PR 17-OCT-1997; 97US-0062287P.
PR 17-OCT-1997; 97US-0063755P.
PR 24-OCT-1997; 97US-0062814P.
PR 24-OCT-1997; 97US-0062816P.
PR 24-OCT-1997; 97US-0063045P.
PR 24-OCT-1997; 97US-0063082P.
PR 24-OCT-1997; 97US-0063127P.
PR 27-OCT-1997; 97US-0063327P.
PR 27-OCT-1997; 97US-0063329P.
PR 28-OCT-1997; 97US-0063550P.
PR 28-OCT-1997; 97US-0063561P.
PR 29-OCT-1997; 97US-0063704P.
PR 29-OCT-1997; 97US-0063733P.
PR 29-OCT-1997; 97US-0063735P.
PR 29-OCT-1997; 97US-0063738P.
PR 03-NOV-1997; 97US-0064248P.
PR 03-NOV-1997; 97US-0064809P.
PR 12-NOV-1997; 97US-0065186P.
PR 17-NOV-1997; 97US-0065846P.
PR 21-NOV-1997; 97US-0066364P.
PR 24-NOV-1997; 97US-0066453P.
PR 24-NOV-1997; 97US-0066511P.
PR 24-NOV-1997; 97US-0066700P.
PR 11-DEC-1997; 97US-0069212P.
PR 11-DEC-1997; 97US-0069278P.
PR 16-DEC-1997; 97US-0069344P.
PR 23-JAN-1998; 98US-0072320P.
PR 04-FEB-1998; 98US-0073612P.
PR 09-FEB-1998; 98US-0074086P.
PR 09-FEB-1998; 98US-0074092P.
PR 12-MAR-1998; 98US-0077791P.
PR 20-MAR-1998; 98US-0078910P.
PR 25-MAR-1998; 98US-0079294P.
PR 27-MAR-1998; 98US-0079663P.
PR 27-MAR-1998; 98US-0079782P.
PR 31-MAR-1998; 98US-0080165P.
PR 12-JUN-1998; 98WO-US012456.
PR 14-JUL-1998; 98WO-US012452.
PR 28-AUG-1998; 98WO-US017888.
PR 10-SEP-1998; 98WO-US018824.

PR 14-SEP-1998; 98WO-US019093.
PR 14-SEP-1998; 98WO-US019094.
PR 14-SEP-1998; 98WO-US019177.
PR 16-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98WO-US019437.
PR 07-OCT-1998; 98WO-US021141.
PR 29-OCT-1998; 98WO-US022991.
PR 29-OCT-1998; 98WO-US022992.
PR 20-NOV-1998; 98WO-US024855.
PR 01-DEC-1998; 98WO-US025108.
PR 05-JAN-1999; 99WO-US000106.
PR 08-MAR-1999; 99WO-US005028.
PR 10-MAR-1999; 99WO-US005190.
PR 20-APR-1999; 99WO-US008615.
PR 14-MAY-1999; 99WO-US010733.
PR 02-JUN-1999; 99WO-US012252.
PR 01-SEP-1999; 99WO-US020111.
PR 08-SEP-1999; 99WO-US020594.
PR 13-SEP-1999; 99WO-US020944.
PR 15-SEP-1999; 99WO-US021547.
PR 05-OCT-1999; 99WO-US023089.
PR 29-NOV-1999; 99WO-US028214.
PR 30-NOV-1999; 99WO-US028313.
PR 30-NOV-1999; 99WO-US028409.
PR 01-DEC-1999; 99WO-US028301.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.
PR 02-DEC-1999; 99WO-US028564.
PR 02-DEC-1999; 99WO-US028565.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 20-DEC-1999; 99WO-US030999.
PR 30-DEC-1999; 99WO-US031243.
PR 30-DEC-1999; 99WO-US031274.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000277.
PR 11-FEB-2000; 2000WO-US003376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 18-FEB-2000; 2000WO-US004342.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US005004.
PR 01-MAR-2000; 2000WO-US005601.
PR 02-MAR-2000; 2000WO-US005746.
XX
XX (GENTH) GENENTECH INC.
PA
PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerlitsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2003-352836/33.
DR P-PSDB; ABU80996.
DR
XX
PT New isolated PRO polypeptide useful for treating diabetes, rheumatoid
PT arthritis, sports injuries, obesity, hearing loss in mammals, stroke, or
PT heart attack.
XX
XX
PS Claim 2; Fig 253; 643pp; English.
XX
XX The present invention relates to the isolation of novel human PRO
CC polypeptides, and the polynucleotide sequences encoding them. The PRO
CC polypeptides are secreted and transmembrane proteins. The PRO
CC polypeptides and polynucleotides are useful for preparing a medicament
CC useful in the treatment of diabetes, bone and/or cartilage disorders
CC (e.g. rheumatoid arthritis, sports injuries, osteoarthritis), obesity,
CC hyper- or hypo-insulinaemia, hearing loss, and coagulation disorders
CC (e.g. stroke, heart attack). Anti-PRO antibodies are useful in diagnostic
CC assays for PRO, by detecting its expression in specific cells, tissues or
CC serum, and for affinity purification of PRO from recombinant cell culture
CC or natural sources. ACA65994-ACA67268 represent cDNA sequences encoding

the human PRO polypeptides of the invention. Note: The sequence data for this patent was obtained in electronic format directly from the USPTO web site at seqdata.uspto.gov/psipdidentry.html

Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

Query Match 9.0%; Score 178.2; DB 7; Length 2016;
Best Local Similarity 54.5%; Pred. No. 1.2e-33;
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

```
QY 76 AGCTCTGCCAAGCACTTTGATGAGGCTTGGCAACGCCCGCTTACCAAGATCGAAGGCGC 135
DB 237 AACCTTCCCTCTTGAGCTTCTCTGAGGCGTGGGCAAGTTCTGCTTACCAAGAGGCGC 296
QY 136 CGTCAAGAAGGTGGCCGCCGCTCCATCTGGGACACGTACTGCCACCTGAGCCATC 195
DB 297 CTGGGACCAAGACGGGAAAGGCGCTTACATCTGGGACGCTTTCACACAGAGTGGAAAGG 356
QY 196 GCGCACCAACG3--CGCCAAAGGCGATGAGCTTGGCATCTACCAACCGCTACGATGA 252
DB 357 GAAAGTGCTGGGAATGAGACGGCAGATGAGCTGTGACGCGTACTACAAAGTCCAGGA 416
QY 253 GGAATTGATCTCTTGACCAAGTACGGGCAAAAGGCTTACCGCTTCTCTGTGATGATC 312
DB 417 GGACATCATCTTGTGAGAGGAACTGCACGTCACCACTACCGATTTCTCTGTGAGCC 476
QY 313 GGGATTCATTCCTTCGCGCGGCGAGGCTGATCCGTCACAGAGAGGGAATTGATTTA 372
DB 477 CCGGCTCTCGGCCCAAGGATCCGAGCCGAGCAGGTGAACAAAGGGAATGAAATCTA 536
QY 373 CAGCAAATGATGATGAGCGCTTGTGAGGCGGGGTATCAAGCCTTGGGTGACTTTGTACA 432
DB 537 CAGTATCTTATGATGAGTCCCTTCTGACGACAAACATCACTCCATCGTACCTTGACCA 596
QY 433 CTGGGATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTCAAAGTGAAGAGT 492
DB 597 CTGGGATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTCAAAGTGAAGAGT 656
QY 493 CCAAGCTGACCTTTGAGCGGTATGCGAGGTTGTGCTTTGAAGCTTTTGGGAGCCGAGTCA 552
DB 657 CAA---CTACTTCAGAGACTACGCCAACCTGTGCTTTGAGGCTTTGGGAGCCGTGTGA 713
QY 553 GAATGGATCAACATCAAGAMCCCTGATTCAGGCCATCTATGATATGACACCGGCGAG 612
DB 714 GCACTGATCACTTCAATGATCTCTGCGCAATGACGAAAGGCTATGAGACGGGCCA 773
QY 613 CAAGCCCGCGGAGGAGGACATTAACAAGCACTCCACGAGGCAACACTGCACTGA 672
DB 774 CCAATGCGCGCGGCTGAAGCTCCG-----GGCACCG 806
QY 673 GCCGTGGCTGCTGGAAGGCCCAAGATCATGAGCCATGCCCGCGCGGTCTACAG 732
DB 807 CCTGTACAAAGCAGCAACACATCAATTAAGGCCCAAGCCAAACCTGGCATTTCTTATA 866
QY 733 CAGGGAATTTCGCCCTCGCAAAAGGCGAGATCGGATCTCGCTCAACGCGCATACTA 792
DB 867 CACCACTGGCGCAGAGAGAGAGTCTGTGGGAATTCACCTGAATGTGACTGGGG 926
QY 793 TGAAGCTTGAGACAGCATGAGCTCGGACAAAGAGGCTGTGAGCGACGATGGAATT 852
DB 927 GGAACCTGTGAGACATTTAGTAACCCCAAGAGACTTGAAGGCTGCCAGAGATCACTA 986
QY 853 TCACATGCTGCTGTTTGGCAATCCCATCT 881
DB 987 CTGTCTGGGCTGTTTGGCAACCCCATTT 1015
```

Search completed: March 28, 2004, 17:03:07
Job time : 801.658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 20:52:22 ; Search time 144.443 Seconds
(without alignments)
7649.443 Million cell updates/sec

Title: US-10-026-140-1
Perfect score: 1991
Sequence: 1 agcagcgtccgacacacagcagc.....tttataaaaaaaaaaaaaa 1991

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	7.5	1521	US-09-914-841A-14	Sequence 14, Appl
2	139.4	7.0	1467	US-09-489-039A-2386	Sequence 2386, Ap
3	130.8	6.6	1317	US-09-134-078-2	Sequence 2, Appl
4	128.2	6.4	3163	US-09-344-510B-6	Sequence 6, Appl
5	128.2	6.4	3435	US-09-344-510B-7	Sequence 7, Appl
6	127.8	6.4	1650	US-09-344-510B-9	Sequence 9, Appl
7	127.8	6.4	5032	US-09-344-510B-8	Sequence 8, Appl
8	126.2	6.3	1287	US-09-914-841A-2	Sequence 2, Appl
9	124.8	6.3	1542	US-09-122-230-8	Sequence 8, Appl
10	124.8	6.3	1909	US-09-122-230-6	Sequence 6, Appl
11	118.6	6.0	1828	US-08-688-988-9	Sequence 9, Appl
12	111.4	5.6	1545	US-08-400-275-17	Sequence 17, Appl
13	104.8	5.3	1458	US-09-134-000C-2890	Sequence 2890, Ap
14	101.4	5.1	1437	US-09-107-532A-1268	Sequence 1268, Ap
15	100	5.0	7218	US-08-232-463-14	Sequence 14, Appl
16	95.6	4.8	1410	US-09-489-039A-1053	Sequence 1053, Ap
17	93.8	4.7	1407	US-08-688-988-7	Sequence 7, Appl
18	90.4	4.5	2435	US-09-306-593-1	Sequence 1, Appl
19	88.6	4.5	1931	US-09-431-470-1	Sequence 1, Appl
20	88.6	4.4	1931	US-09-431-470-3	Sequence 3, Appl
21	87.6	4.4	454	US-09-615-192A-166	Sequence 166, App
22	87.6	4.4	454	US-09-169-789-166	Sequence 166, App
23	85.2	4.3	457	US-08-975-316-79	Sequence 79, Appl
24	85.2	4.3	457	US-09-615-192A-79	Sequence 79, Appl
25	85.2	4.3	457	US-09-169-789-79	Sequence 79, Appl
26	85.2	4.3	470	US-09-615-192A-141	Sequence 141, App
27	85.2	4.3	470	US-09-169-789-141	Sequence 141, App

28	84.8	4.3	3460	US-09-344-510B-10	Sequence 10, Appl
29	84.2	4.2	1404	US-09-489-039A-4718	Sequence 4718, Ap
30	84.2	4.2	1455	US-09-489-039A-4800	Sequence 4800, Ap
31	79.4	4.0	1443	US-09-489-039A-1986	Sequence 1986, Ap
32	74.8	3.8	1635	US-08-961-527-74	Sequence 74, Appl
33	71	3.6	1428	US-09-489-039A-2983	Sequence 2983, Ap
34	69.4	3.5	1479	US-09-134-000C-289	Sequence 289, App
35	68.4	3.4	286	US-09-113-294A-7321	Sequence 7321, Ap
36	68	3.4	3615	US-08-920-812-17	Sequence 17, Appl
37	68	3.4	3615	US-08-920-827-17	Sequence 17, Appl
38	68	3.4	3615	US-08-921-177-17	Sequence 17, Appl
39	68	3.4	3615	US-08-962-577C-17	Sequence 17, Appl
40	68	3.4	3615	US-08-920-828-17	Sequence 17, Appl
41	63	3.2	11303	US-08-961-527-115	Sequence 115, Appl
42	62.6	3.1	15231	US-09-128-155-16	Sequence 16, Appl
43	62.6	3.1	176373	US-09-128-155-17	Sequence 17, Appl
44	61	3.1	1359	US-09-019-095A-7	Sequence 7, Appl
45	58	2.9	494	US-08-332-766A-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-09-914-841A-14
Sequence 14, Application US/09914841A
Patent No: 6645750
GENERAL INFORMATION:
APPLICANT: Amamo Enzyme Inc.
TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE
FILE REFERENCE: 066072
CURRENT APPLICATION NUMBER: US/09/914, 841A
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: P. Hei. 11-056299
PRIOR FILING DATE: 1999-03-04
PRIOR APPLICATION NUMBER: PCT/JP00/01242
PRIOR FILING DATE: 2000-03-02
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1521
TYPE: DNA
ORGANISM: Camellia var. sinensis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1521)
OTHER INFORMATION:
NAME/KEY: mat_peptide
LOCATION: (235)..()
OTHER INFORMATION:
US-09-914-841A-14
Query Match 7.5%; Score 150; DB 4; Length 1521;
Best Local Similarity 52.2%; Pred. No. 1.5e-28;
Matches 455; Conservative 1; Mismatches 401; Indels 15; Gaps 5;
QY 81 TGGCCACGACTTGGATGGGCTTGGCAAGCCCGCTTACCAAGCGAGGCGCTCA 140
DB 116 TTCTGATGTTTGTCTTGGAGCGCTTCTTCCCTTACCGATTGAAGGTGCCA 175
QY 141 AAGAAGTGGCCGCGCCGCTTCATCTGGAGACAGTACTGCCACCTGAGCCATCCGCA 200
DB 176 AGAAGGTGGGAAGGCCCAATTTTGGGATACCTTCACTCATGAGTTTCCAGTAA 225
QY 201 --CAAGCGCGCAAGCGGATGGCTTGGCATCACTACCAAGCGCTACGATGAGACT 257
DB 236 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 295
QY 258 TTATCTCTTGAACCAAGTACGCGCAAGCGCTTCTTCTTCTTCTTCTTCTTCTT 317
DB 296 TGAAGTGTGGAAGTTTATGAGCTGATGATGATGATGATGATGATGATGATGATG 355
QY 318 TCATTCCTCTGCGGCGAGGCTGATCCGTCAACAGAGGGAATTGATTTACAGCA 377

Db 356 TATTACCTGGGGGAAAGCTTAGCGAGAGTGAACAAGGAAGTATCCGCTTCTACACA 415
Qy 378 AACTGATTAACGCCCTGTGTAGGCGGGGATATCAAGCTTGGTGACTTTGTAACAAGG 437
Db 416 ATGTCAATCAATATACCTTTATATGAAAGGTATACAACCTTTTATTAACAATCTTCACTGG 475
Qy 438 ATTCGCTCAGGCGCTTCAAGATCGCTATGAGAGGCTGGCTCAACGTGGAAGGCTCCAGC 497
Db 476 ATCTTCCCAAGCCCTAGAGATGATATGAGAGCTTTTAAAGC---CCACATTTGGA 532
Qy 498 TGAATCTTGAAGCGATATGAGAGGTTGTGCTTTGAACGTTTGGGAGACGAGTCCAACT 557
Db 533 ACGATTTCCGGGATTTTGAAGAGCTGTGCTTCAAGAGATTGTGACCGAGTTAAACAT 592
Qy 558 GGATCAACATCAACGAMCCCTGATTCAGAGCCATCTATGATATATCCACCGGAGCAACG 617
Db 593 GGATCAACATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 652
Qy 618 CCCCAGGAGAGAGCAATTAACAAGCACTC---CACGAGGAGCAACACTGCACTGAGC 674
Db 653 CACCGGGCGGTTGTTGCGCTTTATGGAATTTTGGCCATAAGGAAATCTGGAGCTGAGC 712
Qy 675 CGTGGCTGCTGGAAGGCGCAATGATGAGCCATGCCCCGCGTGGCCGCTTACAGCA 734
Db 713 CCTATATAGTAAACCAATTTGCTTCTCATGCTGCTGCTGTAAGTATACAGG 772
Qy 735 GGGACCTTTCGCCCCGCAAAAAGGCGCAATGCGGATCTCGTCAACGCGCACTATG 794
Db 773 AGAATATACAGGATATCAAAAAGGCGCAATGCGGATGACTGATGATATGATG 832
Qy 795 AGCCCTGGAGACGAATGAGCCTCGGAGCAAGAGGCTGCTGAGGAGAGATGATATTC 854
Db 833 TTCCCT---ACTCCATTTGAAAGCGCAAGAGATGAGCAACAAGCCCTTATTTCA 889
Qy 855 ACATTGCTGCTTGGCCATCCATCTTCTTGAAGAAGATATCCAGAGCATGAGA 914
Db 890 TGTATGATGATGTTATGATGAGCATTAAGCTT---TGTGATATATCAAAAAGCATGCGTA 946
Qy 915 AGCAGCTGGGCGAGAGCTTCCAGCCCTCACT 946
Db 947 GACTCGTTGGTAAAGGTTACCAAGTTCACT 978

RESULT 2
US-09-489-039A-2386
; Sequence 2386, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIORITY FILING DATE: 2000-01-27
; PRIORITY FILING DATE: US 60/117,747
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2386
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2386

Query Match 7.0%; Score 139.4; DB 4; Length 1467;
Best Local Similarity 56.1%; Pred. No. 7.5e-26;

Matches 307; Conservative 1; Mismatches 227; Indels 12; Gaps 2;

Qy 77 GCTTCCCAACGATTTGAATGGGGCTTGGCAAGGCGCTTACCAAGATGAGGCGCC 136
Db 103 GCTTTCCGACAAATTTCTTATGGGGTGGCGCTACCGAGGATTCAGGTTGAGGAGGA 162
Qy 137 GTCAAGAAGGTGGCGCGCGCTCATCTGGGACACGTAAGTCCACCTGAGACCATCG 196

Db 163 CACACGCCAGCGGTAAGAGACCTCAATCTGGGATATCTACTCTCATCTCCGGGATCC 222
Qy 197 GCACCAACCGCGCAACCGGATATGAGGCTTGGCATCATCAACCGCTACGATGAGAC 256
Db 223 ACCTTGAAGGACACCGGCGATATGCGGTGACCATATCAACGTTTTCGGAGAAC 282
Qy 257 TTTATCTCTTGAACAGTACGCGGCAAGGCTTACCGCTTCTCTTGTGCTGCTGCGG 316
Db 283 GTGGCGTTAATGCGCGATGAGGCTTGCACAAAGCATATGATTTCTCATCTCTGCCCCG 342
Qy 317 ATCATTTCCCTCGGCGGAGCTGATCCGTCACAGAGGAGGAATGATTTTACAGC 376
Db 343 CTGCTGCG-----CGCCGGCGCGGATAGATGATGAGCGGAGTCCAGTTTACAGC 396
Qy 377 AAATGATTAACGCGCTTGTAGGCGGGATATCAAGCTTGGTGAATTTGTAACAATGG 436
Db 397 GATCTGATGACGAACGCTGCGGCAACATATGAGCCGATGATCAACCTCTATCACTGG 456
Qy 437 GATCTGCTGAGGCGCTTTCAGATCGCTATGAGAGCTGCTCAACGTGGAAGGTTCCAG 496
Db 457 GATCTGCGAGGCGCTTCCAGG-----ATGAGGCGGCTGGGAAAGCGCACTACCGCC 510
Qy 497 CTGACTTTGAGCGGATGCGAGGTTGTGCTTGAACGTTTGGGAGCGAGTCCAGAAC 556
Db 511 GAGGCGCTTGGCGAGTACCGCGCGCTGTGCTATGCGCGCTTGTTCAGGGTGAAGCTG 570
Qy 557 TGTATCAATCAACGAMCCCTGATTCAGGCGCATATGATATGATGATGATGATGATGAT 616
Db 571 TGGGCTTACCTTCAACGAACCATGTTCTTATGAGGACAGGCTAATTAACGCGCTCAT 630
Qy 617 GCCCGCG 623
Db 631 CCCCCG 637

RESULT 3
US-09-134-078-2
; Sequence 2, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1317 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 1..1314
 US-09-134-078-2

Query Match 6.4%; Score 130.8; DB 4; Length 1317;
 Best Local Similarity 56.4%; Pred. No. 1.1e-23;
 Matches 309; Conservative 1; Mismatches 223; Indels 15; Gaps 3;

```

QY 77 GCTCTGCCCCAAGACATTTGAAATGGGGCTTGGCAACGGCCGCTTACCGCATCGAAGGGCC 136
DB 16 GATTTTCCAAAAGATTTTATCTTCGGAACGGCTACCGACATACCGAATTGAAGTGCA 75
QY 137 GTCAGAGAGAGTGGCCGCGCCGCTCATCTGGAGACAGTACTGCCACTGAGCCATCG 196
DB 76 GCAGAGAGAGATGGCAGAGGGCCATCAATTGGAGTGTCTTTCACACAGCCCTGGCAA 135
QY 197 CGCAGCAACGGCCGCAACGGCGATGGCTTGGCATCTACACCGCTACGATGAGAC 256
DB 136 ACCCTGAAGGTGACACAGAGAGCGTGGTGACCATTAATACCGATCAAGGAAGAT 195
QY 257 TTGATCTCTTGAACCAAGTACGGCGCAAGAGCCCTTCTCTTGTGTGTGCGCG 316
DB 196 ATCCAGCTGATGAAGAAAGATAGGATTAAGACGCTTACAGTTCTCTCTGCGCCAGA 255
QY 317 ATCATTCCTCCCTCGCGGCGAGCTGATCCGCTCAACAGAGAGGAATTGATTTTACAG 376
DB 256 ATTATGCGCATG--GAGAACATCAACCAAAAGGTGGATTTCTTACAC 306
QY 377 AAACGATTTGACGCCCTGTTGAGCGGGATATCAACGCTTGGTGAATTTTACAGT 436
DB 307 AGACTCGTTGATGAGCTTTTGAAGAAATGATATCATACATTCGTAACACTCTATCACTGG 366
QY 437 GATCTGCTTGAAGCGCTTCAAGATCGCTATGAGAGGCTGCTCAACGCTGAAGAGTCCAG 486
DB 367 GACTTACCTTACCACTTATGAAA--AAGGTGAGTGGCTTAAACC--AGATATAGCG 420
QY 497 CTGACCTTTGAGGATATGCGAGGTTGTGCTTTGAACGTTTGGGAGCGAGTCCAGAAC 556
DB 421 CTCTATTTCAGACATACGCAACGTTTATGTTCAACGAACCTCGTATCGTTGAACAT 480
QY 557 TGGATCACCATCAACGAMCCCTGATTCAGGCACTATGATATGCAACGGCAGCAAC 616
DB 481 TGGATTACACTGAACGAACCATGTGTCTTCTTCTCGGTTATATCAAGGAGAGCAT 540
QY 617 GCGCCGGG 624
DB 541 GCGCCGGG 548

```

RESULT 4

US-09-144-510B-6
 Sequence 6, Application US/09344510B
 Patent No. 6579850
 GENERAL INFORMATION:
 APPLICANT: Nabeshima, Youichi
 Kuroo, Makoto
 Sekine, Susumu
 Iida, Akihito
 TITLE OF INVENTION: No. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1
 NUMBER OF SEQUENCES: 38
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
 STREET: 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: United States

ZIP: 10112-3801
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 1.44 mb, DS, DD
 COMPUTER: Compaq Deskpro EN
 OPERATING SYSTEM: Windows 98
 SOFTWARE: Wordpad
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/344,510B
 FILING DATE: 25-Jun-1999
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP97/04585
 FILING DATE: 12-DEC-1997
 APPLICATION NUMBER: JP 347871
 FILING DATE: 26-DEC-1996
 APPLICATION NUMBER: JP 205815
 FILING DATE: 31-JUL-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Perry, Lawrence S.
 REGISTRATION NUMBER: 31865
 REFERENCE/DOCKET NUMBER: 766.32
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 218-2100
 TELEFAX: (212) 218-2200
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3163
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: human
 IMMEDIATE SOURCE:
 LIBRARY: kidney
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 9..3047
 IDENTIFICATION METHOD: E
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 US-09-144-510B-6

Query Match 6.4%; Score 128.2; DB 4; Length 3163;
 Best Local Similarity 52.1%; Pred. No. 7.3e-23;
 Matches 407; Conservative 0; Mismatches 338; Indels 36; Gaps 4;

```

QY 191 CCATCGGCACCAACGCGCCCAACGGCGATGGCTTGGATCTACCAACCGCTACGAT 250
DB 348 CCGTCGCGGTGACGCCCGCACCGGGGAGCTAGCCAGGACAGCAACACGCTTTC 407
QY 251 GAGGACTTTGATCTTGAACCAAGTACGGCGCAAGAGCCCTACCGCTTCTGTGCTGG 310
DB 408 CGCGACAGGAGCGCGTGGCGAGCTGGGATCACTACCTACCTTCTTCATCTGTGG 467
QY 408 CGCGACAGGAGCGCGTGGCGAGCTGGGAGCTGGGAGCTGGGAGCTGGGAGCTGGGAG 584
DB 311 TCGCGATATTCCTCGCGCGGCGAGCTGATCCGCTCAACGAGAGGAATTGATTT 370
QY 468 GCGGAGTGTCTCCCAATGAGCAGCGGGGGGTCC--CAACCGGAGGGGCGCTGAC 524
DB 371 TACAGCAAACTGATTGAACGCGCTTGGAGCGGGGATATCAACGCTTGGGATTTGAT 430
QY 525 TACCGGCGCTGTGAGAGCGGTGGCGGAGCTGGGAGCTGGGAGCTGGGAGCTGGGAG 584
DB 431 CACTGGGATCTGCTCAGGCGCTTCAAGATCGCTATGAGAGGCTGGCTCAACGTAAGAG 490
QY 585 CACTGGGAGCTGGCCCAAGCGCTGCGAGAGCCTTACGCGGCTGGGCAACCGCGCTTG 644
DB 491 GTCAGCTGACTTTGAGCGGTATGCGAGTGTGCTTTGAACGTTTGGGAGCGAGTC 550
QY 645 GCGGACC--ACTTACGAGATTACGGGAGCTGTCTTCGACCACTTGGCGGCTGAGTTC 701
DB 551 CAGAACTGATCACTCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGGC 610
QY 702 AAGTACTGATCACTCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGGC 761

```

QY 611 AGCAAGCCCCGAGAGAGAGATTAACAAGCACTCCAGGAGCAACTGCACT 670
DB 762 CGCGTGGCCCCCGGATCCGGGGGAGCCCGGCTC----- 797
QY 671 GAGCGCTGCTGCTGGAAAGCCAGATGATGAGCATGCCGCGCTGCGCTTAC 730
DB 798 ---GGGTACTGCTGGTGGCAACCTCTCTGCTGATGCAAGATGTCATCTTAC 854
QY 731 AGCAGGAGATTTCGCCCCCTCGCAAAAGGCGAGATGCGCATTCCTCAACGCGACATAC 790
DB 855 AATACTTCTTCCGCTCCACTAGAGAGGTGATGCTTCTATGCTTCAAGCTTCTAC 914
QY 791 TATGAGCCCTGGAGACAGCATAGGCTCGGCAAGAGAGCTGTAGCGAGATGAA 850
DB 915 ATCAATCTCGAAGATGACGACCAAGCATCAAGATGCAAAATCTC---TGGAC 971
QY 851 TTTCACATTTGCTGTTGGCCATTCCTTCTTGAAGAGACTTTCAGAGCATG 910
DB 972 TTGTACTAGTGTGTTGGCCAAACCGTATTTATTTATGATGATCTATCCGAGAGCATG 1031
QY 911 AAGAAGCACTGGGAGAGAGCTTCAGGCTTCACTCCGCGGACTTTGCCATCCTCAT 970
DB 1032 AAGAATTAACCTTTCACTATCTGCTGATTTACTGATGAGAAAAAGTTCAATCAA 1091
QY 971 G 971
DB 1092 G 1092

RESULT 5

US-09-344-510B-7

Sequence 7, Application US/09344510B

Patent No. 6579850

GENERAL INFORMATION:

APPLICANT: Nabeshima, Youichi

Kuroo, Makoto

Sakine, Susumu

Iida, Akihito

TITLE OF INVENTION: No. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESS: Fitzpatrick, Cella, Harper & Scinto

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: United States

ZIP: 10112-3801

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 1.44 mb, DS, DD

OPERATING SYSTEM: Windows 98

SOFTWARE: Wordpad

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/344,510B

FILING DATE: 25-Jun-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP97/04585

FILING DATE: 12-DEC-1997

APPLICATION NUMBER: JP 347871

FILING DATE: 26-DEC-1996

APPLICATION NUMBER: JP 205815

FILING DATE: 31-JUL-1997

ATTORNEY/AGENT INFORMATION:

NAME: Perry, Lawrence S.

REGISTRATION NUMBER: 31865

REFERENCE/DOCKET NUMBER: 766.32

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 218-2100

TELEFAX: (212) 218-2200

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 3435

TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: human
IMMEDIATE SOURCE:
LIBRARY: kidney
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1655
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-344-510B-7

Query Match 6.4%; Score 128.2; DB 4; Length 3435;
Best Local Similarity 52.1%; Pred. No. 7.5e-23;
Matches 407; Conservative 0; Mismatches 338; Indels 36; Gaps 4;

QY 191 CCATCGGAGACCAAGCGCGCAACGGGATGCTTGGATCACTACCGCTACGAT 250
DB 348 CCGTCGCGCTGACAGCCCGCCACCGGAGCTAGCCAGACGATCAACAGCTTTC 407
QY 251 GAGGACTTTGATCTTGAACCAAGTACGGCGCAAGGCTTACCGCTTCTTGTCTG 310
DB 408 CGGACACGAGAGGCGGTGGCTGAGCTCGGGGTCACTACCTGCTTCTGATCTGTG 467
QY 311 TCGCGATCTTTCCTCGGCGGAGCTGATCCGCTCAACGAGAGGAATTGATT 370
DB 468 GCGGAGTGTCTCCCAATGAGCAGCGGCGTCC---CAACCGCAAGGCGCTGCGCTAC 524
QY 371 TACAGCAACTGATTAAGACGCTGTTGAGCGGGGTATACGCTTGGGATGATTGAC 430
DB 525 TACCGGCGCTGTGAGCGGCTGCGGAGCTGAGGCTGACGCTGATCAGCTGTAC 584
QY 431 CACTGGATCTGCTTACGCGCTTACGATTCGCTATGAGCTGCTCAACGTGAAGAG 490
DB 585 CACTGGAGCTGCTCCCAATGAGCAGCGGCTGAGAGCGCTACGCGGCTGAGGCAACCGCGCTG 644
QY 491 GTCAGCTGATCTTGAAGGATATGAGGATGCTTGTGTTGAACGTTTGGGACGAGTC 550
DB 645 GCGGACCT---ACTTCAAGGATTAAGCGAGCTTGTGCTTCCGCTTGGCGCTGACGTC 701
QY 551 CAGAACTGATCAACATCAACGAGCCCTGATTAAGGCTATGATGATGATGATGATG 610
DB 702 AAGTACTGATCAACATCAACGAGCCCTGATGATGATGATGATGATGATGATG 761
QY 611 AGCAAGCCCCGAGAGAGAGATTAACAAGCACTCAACGAGGCAACTGCCACT 670
DB 762 CGCTGCGCCCCCGGCAATCCGGGCAAGCCCGGCTC----- 797
QY 671 GAGCGTGGCTGCTGGAAAGGCCAGATCATGAGCATGCCGCGCTGCGCTTAC 730
DB 798 ---GGGTACTGCTGGGCGCAACCTCTCTGCTGATGCAAAAGTCTGGGATCTTAC 854
QY 731 AGCAGGAGATTTCGCCCCCTCGCAAAAGGCGAGATGCGCATCTGCTCAACGCGCATAC 790
DB 855 AATACTTCTTCCGCTCCACTAGGAGGTGAGGTGATGATGATGATGATGATGATG 914
QY 791 TATGAGCCCTGGAGACAGCATAGGCTGGGCAAGAGAGCTGTGAGCGAGATGAA 850
DB 915 ATCAATCTCGAAGATGACGACCAAGCATCAAGATGCAAAATCTC---TGGAC 971
QY 851 TTTCACATTTGCTGTTGGCCATTCCTTGAAGAGACTTTCAGAGCATG 910
DB 972 TTGTACTAGTGTGTTGGCCAAACCGTATTTATTTATGATGATGATGATGATG 1031
QY 911 AAGAAGCACTGGGAGAGGCTTTCAGGCTTCCGCGGAGCTTTCATCTCAT 970
DB 1032 AAGAATTAACCTTTCACTATCTGCTGATTTACTGATGATGAGAAAAAGTTATCAA 1091
QY 971 G 971
DB 1092 G 1092

RESULT 6
US-09-344-510B-9
Sequence 9, Application US/09344510B
Patent No. 6579850
GENERAL INFORMATION:
APPLICANT: Nabeshima, Youichi
Kuroo, Makoto
Sekine, Susumu
Iida, Akihito
TITLE OF INVENTION: NO. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10112-3801
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 1.44 mb, DS, DD
COMPUTER: Compaq Deskpro EN
OPERATING SYSTEM: Windows 98
SOFTWARE: Wordpad
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/344,510B
FILING DATE: 25-Jun-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/04585
FILING DATE: 12-DEC-1997
APPLICATION NUMBER: JP 347871
FILING DATE: 26-DEC-1996
APPLICATION NUMBER: JP 205815
FILING DATE: 31-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Perry, Lawrence S.
REGISTRATION NUMBER: 31865
REFERENCE/DOCKET NUMBER: 766.32
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 218-2100
TELEFAX: (212) 218-2200
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: mouse
IMMEDIATE SOURCE:
LIBRARY: kidney
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1650
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-344-510B-9
Query Match 6.4%; Score 127.8; DB 4; Length 1650;
Best Local Similarity 53.1%; Pred. No. 7e-23;
Matches 388; Conservative 0; Mismatches 307; Indels 36; Gaps 4;

Db 466 GCGGGGTGCTCCCAATGACGCGGGGCACTCC---CAACCGGAGGGGCTGCGCTAC 522
Qy 371 TACAGCAACTGATTTAGCGCCCTTTGAGCGGGGATACGCTTGAGTCTTTTAC 430
Db 523 TACCGGGGCTGTGGAGCGCGCTGCGGAGCTGGCTGAGCGCGTTCACCTTTAC 582
Qy 431 CACTGGATCTGCTCAGCGCTTACAGATCGCTATGAGAGGCTGAGTCAAGTGAAG 490
Db 583 CATGGGACCTGCGACAGCGCTGCGAGACCTATGCGGATGGGCAATCGGCCCTG 642
Qy 491 GTCCAGCTGACCTTTGACGGATATGCGAGTGTGCTTTGAAGCTTTGGGAGCCAGT 550
Db 643 GCGGACC---ATTACAGGATTTATGCGAGCTGTGCTTCGCGCACTTCGGTGGTCA 699
Qy 551 CAGAACTGATCACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCCACCGC 610
Db 700 AAGTACTGATACACATTTGACCAACCCCTACGTGTGCTTGGCAGCGGTATGCCACCGG 759
Qy 611 AGCAAGCGCCCGGAGAGACGATTAACAGCACTCCACGAGGCAACATGCCACT 670
Db 760 GCGCTGGCCCCGGCGTGAAGGGGAGCTCCAGGCTCGG----- 797
Qy 671 GAGCGGTGCTCGCTGAAAAGGCCAGATCATGAGCCATCCGCGCGCTGCGCTAC 730
Db 798 -----GTACCTGGTTGCCCAACACTATTGCTGATGCAAAAGCTGGCATCTCTAC 852
Qy 731 AGCAGGACTTTTGCCCTTCGCAAAAGGCCAGATCGGCATCTGCTCAACGCGACTAC 790
Db 853 AACACCTCTTTCGCGCCACACAGGAGGCGGGGTGCTATGCTTAAGCTCCCATTTG 912
Qy 791 TATGAGCCCTGGGACAGCAATGAGCTCGGGAACAGAGGCTGTGAGCGAGATGAA 850
Db 913 ATCAATCTTGAAAGATGACTGATATATACAGAAATGCCAAGTCTC---TTGAC 969
Qy 851 TTTCATTGAGCTGTGTTGCAATCCCATCTTTTGAAGAGACTATCCAGAGAGATG 910
Db 970 TTGTGCTAGAGCTGTGTTGCAAAACCATATTATTATGAGCGCATCCAGAGATG 1029
Qy 911 AAGAGCAGCT 921
Db 1030 AAGAACCACT 1040
RESULT 7
US-09-344-510B-8
Sequence 8, Application US/09344510B
Patent No. 6579850
GENERAL INFORMATION:
APPLICANT: Nabeshima, Youichi
Kuroo, Makoto
Sekine, Susumu
Iida, Akihito
TITLE OF INVENTION: NO. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10112-3801
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 1.44 mb, DS, DD
COMPUTER: Compaq Deskpro EN
OPERATING SYSTEM: Windows 98
SOFTWARE: Wordpad
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/344,510B
FILING DATE: 25-Jun-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/04585
FILING DATE: 12-DEC-1997

```

; APPLICATION NUMBER: JP 347871
; FILING DATE: 26-DEC-1996
; APPLICATION NUMBER: JP 205815
; FILING DATE: 31-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Perry, Lawrence S.
; REGISTRATION NUMBER: 31865
; REFERENCE/DOCKET NUMBER: 766.32
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 218-2100
; TELEFAX: (212) 218-2200
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5032
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: mouse
; IMMEDIATE SOURCE:
; LIBRARY: kidney
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9..3060
; IDENTIFICATION METHOD: E
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-344-510B-8

Query Match      6.4%; Score 127.8; DB 4; Length 5032;
Best Local Similarity 53.1%; Pred. No. 1.1e-22;
Matches 388; Conservative 0; Mismatches 307; Indels 36; Gaps 4;

191 CCATCGCGACCAACGCGCGCGCGATGCTGGTGGATCACTACCGCTACGAT 250
   |||||
364 CGCTCCCTCCCTCCCTGCTCCCTCACTGAGATGCGCAGAGATGTTACAAACCTTAC 423
   |||||
251 GAGACTTTGATCTTTGACCAAGTACGGCGCAAGGCGCTACCGCTTCCCTTGGTGG 310
   |||||
424 CGGACACAGAGAGGCGCTGCGCAACTGGGGGTCCACCTACCGCTTCCCTTACATG 483
   |||||
311 TCGGGATCATTCCTCCGCGCGCGCGATCGATCCCGTCAAGAGAGGAAATTAGTTT 370
   |||||
484 GCGGGGTGCTCCCAATGCGACCGCGGCGACTCC---CAACCGGAGGCGCTGCGCTAC 540
   |||||
371 TAGACAACTGATTGACGCTTGTGAGCGCGGGGTATCACGCTTGGGTGACTTTGTAC 430
   |||||
541 TACCGGCGCTGCTGAGCGCGCTGCGGAGCTGGGCGTGCGAGCTGTGTACCTGTAC 600
   |||||
431 CACTGGGATGCTGCTGCGCGCTTACGATCGCTATGAGAGGCTGCTCAACGTGGAAG 490
   |||||
601 CATTGGGACCTGCGACAGCGCTGCGAGACCTATGCGGATGGGCGCAATGCGGCCG 660
   |||||
491 GTCCAGCTGACTTTGAGCGGTATGCGAGGTGTGTCTTGAACGTTTGGGGAACGAGTC 550
   |||||
661 GCCGAC---ATTTCAGGATTTATGCGAGCTGTGCTTCGCGCACTTCGATGTCAAGTC 717
   |||||
551 CAGAACTGATCAACATCAACGAMCCTTGATTCAGGCGCATATGATATGCGACCGG 610
   |||||
718 AAGTACTGATCAACATTCATGACACCTTACGCTGCTGCGACGCGGTATGCGACCG 777
   |||||
611 AGCAACGCGCGCGCGAGGAGGAGCTTAAACAGACTTCAACCGAGGCGCAACCTGCCACT 670
   |||||
778 CGCTGCGCGCGCGCGGAGGAGGAGCTCCAGGCTCGG-----815
   |||||
671 GAGCGGTGCTGCTGGAAGGCCAGATCATAGACCATGCGCGCGCGCTGCGCTTAC 730
   |||||
816 -----GTACTGTGTGCGCACAACTTCTTGTGCTCATGCGAAAGTGTGCGATCTTAC 870
   |||||
731 AGCAGGACTTTGCGCGCTCGCAAAAGGCGAGATCGGCTGCTCAACGCGGAGTAC 790
   |||||
871 AACACCTCTTTCGCGCGCACACAGGAGGCGCGGTGTATGCTTAAAGTCTCCCATTTG 930
   |||||
```

```

791 TATGAGCCCTGGGACAGCATATGAGCTCGGACAGAGCGCTGCTGAGCGAGATGAA 850
   |||||
931 ATCAATCTTCAGAAATATGACTATATATATGAGAGATTCAGAGATCTC---TTGAC 987
   |||||
851 TTTCAATTTGCTGCTTTTCCCATTCCTTTTGAAGAAAGCTATCCAGAGAGCATG 910
   |||||
988 TTTGTGTAGGCTGCTTTGCCAAACCCATATTTATGTGCGGACTACCCAGAGATATG 1047
   |||||
911 AAGAGCACT 921
   |||||
1048 AAGAACCACT 1058
   |||||

RESULT 8
US-09-914-841A-2
; Sequence 2, Application US/09914841A
; Patent No. 6645750
; GENERAL INFORMATION:
; APPLICANT: Amano Enzyme Inc.
; TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE
; FILE REFERENCE: 066072
; CURRENT APPLICATION NUMBER: US/09/914,841A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: P. Hei. 11-056299
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: PCT/JP00/01242
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 2
; LENGTH: 1287
; TYPE: DNA
; ORGANISM: Camelilia var. sinensis
US-09-914-841A-2

Query Match      6.3%; Score 126.2; DB 4; Length 1287;
Best Local Similarity 51.5%; Pred. No. 1.6e-22;
Matches 390; Conservative 1; Mismatches 354; Indels 12; Gaps 4;

201 CCAACGCGCGCAACGCGGATGCGCTTGGATCACTACCGCTACCGCTTATGAGAGATTGG 260
   |||||
5 CGAATGTATACACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 64
   |||||
261 ATCTTTGACCAAGTATGAGGCGCAAGGCTTACCGCTTCTCTGTGCTGCGGATCA 320
   |||||
65 AGTGTGTAAGTTATGAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 124
   |||||
125 TACCTCGGAGAACCTTATAGCGAGATGAAAGAGATATCGCTTCAACAAATG 184
   |||||
321 TTCCCTCGCGGCGAGCTGATCCGTCACAGAGAGGAAATTGATTTTACAGCAAC 380
   |||||
381 TGATTTGACGCTTGTGAGCGGGGTATCACGCTTGGGATCTTTGATCACTGGGATC 440
   |||||
185 TCATCAATGACCTTTATGCAAGATATCAACCTTTTATTAACATCTTTCACTGGGATC 244
   |||||
441 TGCCCTGAGCGCTTACAGATGCTATGAGGCTGCTCAACGTTGAAGAGTCAAGCTGG 500
   |||||
245 TTCCCAAGCTTGAAGATGATATGAGGCTTTTAAGC---CCACACTTGTGAACG 301
   |||||
501 ACTTTGACCGGTATGCGAGGTGTGCTTTGAACGTTTGGGAGCGAGTCCAGAACTGGA 560
   |||||
302 ATTTCCGGGATTTTCAAGAGCTGTGCTTCAAGAGATTGAGCGAGTTAAACTTGA 361
   |||||
561 TCACATCAAGCAAGCCCTGATGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 620
   |||||
362 TCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 677
   |||||
621 CGGCGAGAG---CAGACTTAAACAGACTTCAACGAGGCGAACACTGCACTGAGCGCT 734
   |||||
422 CGGCGCTGTGTGCTTTTATGCGATTTTGGCTTAAAGGAAATCTGAGCTGAGCTCT 481
   |||||
678 GGTGCTGGAAGGCGGAGATGATGAGCGGCGGCGGCTGCTTACAGAGG 737
   |||||
```


Db 482 ATATAGTACCACAAATTGCTTTCTCATGCTGCTGTGTAACATAACAGAGA 541
 Qy 738 ACTTGGCCCCCTGGCAAAAGGCGAGATCGGATCTGCTCAACGGGACTACTATGAGC 797
 Db 542 AATATCGAGCATATCAAAAGGCGAGATAGGATTAACATGACTTATTTGATGATTC 601
 Qy 798 CTTGGGACAGCAATGAGCCTTGGGACAAAGAGCTGTGAGCGACGATGGAAATTTACA 857
 Db 602 CCT---ACTCCAAATTCGAAAGCGGACAAAGATGACGACAAAGCCCTTGATTCATGT 658
 Qy 858 TTGGCTGTGTTGCCAATCCATCTTTCTTGAAAGAGATATCCAGAGCATGAAGAGC 917
 Db 659 ATGATGTGTTATTTAGCCATTAAGCTT---TGTGTAATATCCAAAAGCATGCGTAGAC 715
 Qy 918 AGCTGGCGAGAGGCTTCCAGCCCTCACTCCCGCGGA 954
 Db 716 TCCTGTGTAAGAGTTACCAAGCTTCACTAAGAGCA 752

RESULT 9

US-09-122-230-8
 ; Sequence 8, Application US/09122230A
 ; Patent No. 5973228
 ; GENERAL INFORMATION:
 ; APPLICANT: Carlson, et al.
 ; TITLE OF INVENTION: Confiterin Beta Glucosidase cDNA for Modifying Lignin
 ; TITLE OF INVENTION: Composition in Plants
 ; FILE REFERENCE: 50532
 ; CURRENT APPLICATION NUMBER: US/09/122,230A
 ; CURRENT FILING DATE: 1998-07-23
 ; EARLIER APPLICATION NUMBER: U.S. 60/053,566
 ; EARLIER FILING DATE: 1997-07-24
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 8
 ; LENGTH: 1542
 ; TYPE: DNA
 ; ORGANISM: Pinus contorta
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1542)
 ; US-09-122-230-8.

Query Match 6.3%; Score 124.8; DB 2; Length 1542;
 Best Local Similarity 54.4%; Pred. No. 4e-22;
 Matches 299; Conservative 0; Mismatches 242; Indels 9; Gaps 2;

Qy 81 TGGCCCAAGCACTTGAATGGGGCTTTCGCAACGGCCGCTACAGATCGAAGCGCGTCA 140
 Db 89 TCCCTCAGATTTCATGTTTGGGACAGCCTCTTCAAGGTATCAATATGAGAGCAATCC 148
 Qy 141 AAGAAGTGGCCCGGCGCCGCTCCATCTGGGACAGTACTGCGCACTTGAGAGCATCGGCA 200
 Db 149 GAGAAGATGGCAAGGGTCTTAAGCAATGAGGAGCGCTTAAACATATCTCTGTGAATTA 208
 Qy 201 CCAAGCGCGCAACGGGATGTGGCTTGCATCACTACCGCTAGATGAGAGACTTTG 260
 Db 209 AAGATGACAGCAATGAGAGTGGCACTGCAATATTCACAGATATATGAGAGATTCG 268
 Qy 261 ATCTTTGACCAAGTACGGGCGCAAGGCTTACCGCTTCTCTGTGCTGTGCGGATCA 320
 Db 269 AGCTTATGGCTTACCTTGAGACTAGTCTATATATCTCTCATATCTCGGTCTGAAATCC 328
 Qy 321 TTCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGAAATGATTTTACAGCAAC 380
 Db 329 TTCAGAGAGAG---AGGTAAATTAATCATGCGTGGGATTTGAATATTAATATATC 382
 Qy 381 TGATTAAGCGCCCTGTGAGGCGGGATATCAAGCCTTGGGATCTTTGACCACTGGGATC 440
 Db 383 TGATTAAGCGCTTCTTCAAAATGAGATCCAGCCGTTGAGACATTTGTTCAATTTCAATC 442
 Qy 441 TGCCTCAGGCGCTTACAGATGCTATGAGGCTGCTCAACGTGAAAGAGGTCCAGCTGG 500

Db 443 TTCCCAAGCACTTAAGACTCTATGGGGATGCTGAGTCTCTCAATATTA---CG 499
 Qy 501 ACTTGAAGCGGTATGAGAGGTTGCTTTGAACGTTTGGGAGCCAGTCCAGACTGA 560
 Db 500 ACTTCGAAGCTTATGAGAGATTTGCTTCCGGGATTCGTTGACCGGTCAATATTTGGG 559
 Qy 561 TCACATCAACGAMCCCTGATTCAGGCCATATGATATGCAACCGGACGAAACGCC 620
 Db 560 CGACATGAGCGAGCCAAATCTGTTGTGCGGTTGGGATACACCGTCCGAATATTTCCAC 619
 Qy 621 CGGCGAGAG 630
 Db 620 CGACAGGTG 629

RESULT 10

US-09-122-230-6
 ; Sequence 6, Application US/09122230A
 ; Patent No. 5973228
 ; GENERAL INFORMATION:
 ; APPLICANT: Carlson, et al.
 ; TITLE OF INVENTION: Confiterin Beta Glucosidase cDNA for Modifying Lignin
 ; TITLE OF INVENTION: Composition in Plants
 ; FILE REFERENCE: 50532
 ; CURRENT APPLICATION NUMBER: US/09/122,230A
 ; CURRENT FILING DATE: 1998-07-23
 ; EARLIER APPLICATION NUMBER: U.S. 60/053,566
 ; EARLIER FILING DATE: 1997-07-24
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 1909
 ; TYPE: DNA
 ; ORGANISM: Pinus contorta
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (183)..(1724)
 ; US-09-122-230-6.

Query Match 6.3%; Score 124.8; DB 2; Length 1909;
 Best Local Similarity 54.4%; Pred. No. 4.3e-22;
 Matches 299; Conservative 0; Mismatches 242; Indels 9; Gaps 2;

Qy 81 TGGCCCAAGCACTTGAATGGGGCTTTCGCAACGGCCGCTACAGATCGAAGCGCGTCA 140
 Db 271 TCCCTCAGATTTCATGTTTGGGACAGCCTCTTCAAGGTATCAATATGAGAGCAATCC 330
 Qy 141 AAGAAGTGGCCCGGCGCCGCTCCATCTGGGACAGTACTGCGCACTTGAGAGCATCGGCA 200
 Db 331 GAGAAGATGGCAAGGGTCTTAAGCAATGAGGACCGCTTAAACATATCTCTGTGAATTA 390
 Qy 201 CCAAGCGCGCAACGGGATGTGGCTTGCATCACTACCGCTAGATGAGAGACTTTG 260
 Db 391 AAGATGACAGCAATGAGAGTGGCACTGCAACATATTAACATATATGAGAGATTCG 450
 Qy 261 ATCTTTGACCAAGTACGGGCGCAAGGCTTACCGCTTCTCTGTGCTGTGCGGATCA 320
 Db 451 AGCTTATGGCTTACCTTGAGACTAGTCTATATATCTCTCATATCTCGGTCTGAAATCC 510
 Qy 321 TTCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGAAATGATTTTACAGCAAC 380
 Db 511 TTCAGAGAGAG---AGGTAAATTAATCATGCGTGGGATTTGAATATTAATATATC 564
 Qy 381 TGATTAAGCGCCCTGTGAGGCGGGATATCAAGCCTTGGGATCTTTGACCACTGGGATC 440
 Db 565 TGATTAAGCGCTTCTTCAAAATGAGATCCAGCCGTTGAGACATTTGTTCAATTTCAATC 624
 Qy 441 TGCCTCAGGCGCTTACAGATGCTATGAGGCTGCTCAACGTGAAAGAGGTCCAGCTGG 500
 Db 625 TTCCCAAGAGACTTTGAGAGACTCTATGAGGAGATGCTGAGTCTCTCAATATTA---CG 681
 Qy 501 ACTTTGAGCGTATGCGAGGTTGTCTTGAACGTTTGGGAGCGAGTCCAGAACTGA 560

Db 682 ACTCGAAGCTATGACAGATTGCTCCGGGCAATGCGTACCGGTCAATATATGGG 741
Qy 561 TCACCATGAAGAMCCCTGATTCAGGCCATCTATGATATGATCCAGCGGACGACACGCC 620
Db 742 CGACAGTAAAGAGCAAAATCTGTTTGTGCGGTGGATACCGGTGGAAATATTTCCAC 801
Qy 621 CGGCGAGAGG 630
Db 802 CGACGAGGTG 811

RESULT 11
US-08-688-988-9
; Sequence 9, Application US/08688988B
; Patent No. 6096545
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Daniel D.
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
; FILE REFERENCE: PPL96-03
; CURRENT APPLICATION NUMBER: US/08/688,988B
; CURRENT FILING DATE: 1996-07-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1829
; TYPE: DNA
; ORGANISM: Arabidopsis Thaliana
US-08-688-988-9

Query Match 6.0%; Score 118.6; DB 3; Length 1829;
Best Local Similarity 53.6%; Pred. No. 1.6e-20;
Matches 294; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

Qy 79 TCTGCCCAAGACCTTGAATGGGCTTGCAGACCGCCCTACAGATCGAAG---CGC 135
Db 159 TTTCCTGGAAGCTTTTATTTGGACGCGCTACTCGGCATACAGAGTACCAAGGTTGA 218
Qy 136 CGTCAAGAAAGTGGCGCGCGCCGCTCCATCTGGGACACGACTGCCACTGAGCCATC 195
Db 219 TTTATGAACTTGTGTGTGAGACCGCTTATGGACATCTACTGTAGAATATCCAGA 278
Qy 196 GCGCACCAAGCGCGCAAGCGCGATGTGCTTGGATCACTACACCGCTACGATGAGA 255
Db 279 GAGGTGCAATTAACATTAACGCGATGTGCGGTGATTTCTTCATCGTTAAGGAAGA 338
Qy 256 CTTTGATCTTGAACAAAGTACGCGGCAAGGCTTCCCTGCTGTGTGTGCGG 315
Db 339 TATCAACTATATGAAGATCTTAAACACAGACGCTTTAAGATGCTATCGCATGCGCAG 398
Qy 316 GATCATTCCTCGGCGGACGCTGATCCGTCACAGAGGAGAAATTGATTTTACAG 375
Db 399 AATATTTCTCATGGGAGAAAGAGAAAGAGTGAAGTGAAGTGTGCAATTTCCAG 458
Qy 376 CAAACTGATTAAGCCCTGTGAGGCGGAGTATCAGCGCTTGGTACTTTGTACCATG 435
Db 459 CGACCTATTCAGACGATCAATAAAAATGATTAATCTCAATCTGTTTTCACGTG 518
Qy 436 GGATCTGCTCAGGCGCTTACGATCGCTATGAGGAGGCTGCTCAACGTAAGAGGTCCA 495
Db 519 GGAACATTCACAAATTTTGAAGATGATATGCGGCTTTTAAAGC---GAAAGATTTGT 575
Qy 496 GCTGACCTTTGAGCGGATATGAGAGTGTGCTTGAACGTTTGGGAGCCAGATCCAGAA 555
Db 576 GAAAGATTTCCGAGATATGAGATTTGTTTCCAGAAATACGATGGAGAAAGTGAACA 635
Qy 556 CTGATATCAACATGAACGAMCCCTGATTCAGGCACTATGATGATGACCGGACGAA 615
Db 636 TTGATATCACTTCAATAGACATGGGTTTCTCCACGCTGCTATGACGTAGGCAAAA 695
Qy 616 CGCCCGGG 624
Db 696 GGCACCTGG 704

RESULT 12
US-08-400-275-17
; Sequence 17, Application US/08400275
; Patent No. 5668295
; GENERAL INFORMATION:
; APPLICANT: Wahab, Samir Z.
; APPLICANT: Malik, Vedpal S.
; TITLE OF INVENTION: PUTRESCINE N-METHYLTRANSFERASE,
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING PUTRESCINE
; TITLE OF INVENTION: N-METHYLTRANSFERASE, AND TRANSGENIC TOBACCO PLANTS WITH
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Ave. of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,275
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,681
; FILING DATE:
; APPLICATION NUMBER: US 07/613,160
; FILING DATE: 14-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
; REFERENCE/DOCKET NUMBER: PM-1696
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1545 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-400-275-17

Query Match 5.6%; Score 111.4; DB 1; Length 1545;
Best Local Similarity 49.9%; Pred. No. 1e-18;
Matches 364; Conservative 0; Mismatches 356; Indels 9; Gaps 3;

Qy 75 TAGCTGCGCAACGACTTTGAATGGGCTTCCGAACGCGCGCTACAGATGAGAGCG 134
Db 2 TGCTCTTTTCTTAATTTCTTTGTGCGACAGCCTTCTATTTACCAATGAAAGAG 61
Qy 135 CCGTCAAGAAAGTGGCGCGCGCCGCTCCATCTGGGACAGTACTGCACTGAGACAT 194
Db 62 CTTTCCAGATGATGGAAAGGCTCAGCAACTGGGCGTTTTTACCAATGAAAGGTGTC 121
Qy 195 CGGCGACCAAGCGCGCCAGCGAGATGTGCTTGCGATCACTACGACCGTACAGTAGAG 254
Db 122 ATGTAGGATGGAACCAAGAGAGATGTGCTGTGATCACTACATCGTTATTTGAGAG 181
Qy 255 ACTTGATCTCTGACCAAGTACGCGCAAGGCTTACCGCTTCTGCTGTGCGTGGCG 314
Db 182 ACATCAACTCATGACAGATATGGGTGTGAATGCTTGTGTTCTATCTATGAGGC-- 239

QY 315 GATCATTCCTCGGCGGAGGCTGAGTCCGTCACAGAGAGGAATTGATTTTACA 374
| | | | |
Db 240 -AAGAAATTCGCCAAGGGAATATTGGAGAGATTATATGCGGAAATTGACGACTACA 298
| | | | |
QY 375 GCAAACTGATTTGACGCTGTTGAGGCGGGGTATCAGGCTTGGGTGACTTTGTACACT 434
| | | | |
Db 299 GTAAAGCTCATATGATGACTCTCAAGAAAGGATCCAGCCGTTTGTCACTTAACACTT 358
| | | | |
QY 435 GGGATCGCTCGGCGGCTTCAGCATGCTATGAGAGGCTCAACGTCGAAAGAGTCC 494
| | | | |
Db 359 TTACACATACCAAGAACTTTGAGACAGATATGCTGTTGCTTAA--GTTACAGATAC 415
| | | | |
QY 495 AGCTGACCTTTCGCGGTATGCGAGGTGTGCTTTGAAAGCTTTTGGGACGAGTCCAGA 554
| | | | |
Db 416 GGGATGATTTTACGCTATTTGGCAAACTATGCTTCAAACTGCGGAGATAGAGTTAAT 475
| | | | |
QY 555 ACTGATTCACATCAACGACCCCTGATTCAGGCTATGATGATGATGACCGGACGA 614
| | | | |
Db 476 ACTGAGGTAAAGATGAGGCTTAACTTCGTCGCTATGCTATGAGATGGGACTT 535
| | | | |
QY 615 AGCGCCGCGGACAGAGC--AGCATTAACAAGCACTCCAGGCGCAACTGCGCACTG 671
| | | | |
Db 536 GCGCTTCAATCATGCTGCTGCTATATTTGGAAATGTAGCTGCGGATTCAGAAAGG 595
| | | | |
QY 672 AGCGGTGCTGCTGGAAGGCCAGATCATGAGCCATGCGCGCGCTGCGCTTACA 721
| | | | |
Db 596 AACCTTTCATTTGAGCTCAATATATGATCTATCTATGAGATGCTGTGAGCATTTACC 655
| | | | |
QY 732 GCGAGGACTTTCGCTCCGAAAAGGCGCATGCGCATCTGCTCAACGGGACTACT 791
| | | | |
Db 656 GACACGATATCAAGAAAGTCAAGAGCATGATGCGCATTTACTATGCTTGGATGCT 715
| | | | |
QY 792 ATGAGCCCT 800
| | | | |
Db 716 ATGACCGCT 724
| | | | |

RESULT 13

US-09-134-000C-2890
Sequence 2890, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2890
LENGTH: 1458
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-134-000C-2890

Query Match 5.3%; Score 104.8; DB 4; Length 1458;
Best Local Similarity 53.4%; Pred. No. 4; Be-17;
Matches 267; Conservative 1; Mismatches 223; Indels 9; Gaps 2;

QY 76 AGCTGCGCAAGAGCTTTGATGAGGCTTCGCAAGCGCGCTACAGATCGAAGGCGC 135
| | | | |
Db 45 AGGATTTCCGCAAGATTTTATGAGGCTCTGCTTCGCGAGCTTATCAAGTGAAGGTGC 104
| | | | |
QY 136 CGTCAAGAGAGTGGCGCGCGCGCTGCTCATCTGGGACACGTAAGTCCACTGAGGACTC 195
| | | | |
Db 105 TTGGAAATGAAGAGGCGCAAGCGCATCTGTTGGGATGATTTGTTGGCTTACAGAGAA 164
| | | | |
QY 196 GCGCAACCAAGCGCGCAAGCGCATGCTGCTTGGGATCACTACCAAGCTTACATGAGGA 255
| | | | |
Db 165 AACGTTCAAGAGACAACTGCGATTTAGCGCTGATCATATCATGTTTCAAGAGAA 224
| | | | |

QY 256 CTTCGATCTCTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTTGTGCTGCGG 315
| | | | |
Db 225 TGTGCTTTAATGAACAAACAAAGTTTAAAGATATCGTTCTCGATGATGACGACG 284
| | | | |
QY 316 GATCATTTCCCTCGCGGAGGCTGATCCGTCACAGAGAGGAATTGATTTTACAG 375
| | | | |
Db 285 AATTTTACC-----TGAAGGCGCTGTGAGTAATCAAGCGGAGCTGAAATTTATTC 338
| | | | |
QY 376 CAACGATTTGAGCGGCTTTCGAGGGGATACAGCGCTTGGGTGACTTTTACACTG 435
| | | | |
Db 339 AGATTTAATTTGATGATTTATAGCGGAGAAATTAACCAATGTAACCTTTGATCAATG 398
| | | | |
QY 436 GATGCTGCTCAGGCGCTTCACGATGCTATGAGAGGCTGCTCAACGTGGAAGGTCGA 495
| | | | |
Db 399 GATTTGCCAGCTGTTTTCGCAAAAAGATATGCTGCTG--GATCAAGAAAATTTAT 455
| | | | |
QY 496 GCTGACTTTGAGCGGTATGCGAGTTGCTGCTTTGAAAGCTTTTGGGACCGAGTCCAGA 555
| | | | |
Db 456 TGATGATTTTGTGGCTTATGCGAAATTTTATTTGACGCTTCCGCGCAAGTTCGCTA 515
| | | | |
QY 556 CTGGATCACCATCAAGAAC 575
| | | | |
Db 516 TTGATTTAGTTTAATGAAC 535
| | | | |

RESULT 14

US-09-107-532A-1268
Sequence 1268, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 7310
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCIT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1268:
SEQUENCE CHARACTERISTICS:
LENGTH: 1437 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: circular
TOPOLOGY: circular
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

QY 181 CCACCTGAGGCGATCGGCGACCAACGCGCGCATGATGCTTGCATCACTACCA 240
DB 181 CCACCTGAGGCGATCGGCGACCAACGCGCGCATGATGCTTGCATCACTACCA 240
QY 241 CCGCTACGATGAGGACTTGTATCTTTGACCAAGTACGCGCGCAAGGCTTTC 300
DB 241 CCGCTACGATGAGGACTTGTATCTTTGACCAAGTACGCGCGCAAGGCTTTC 300
QY 301 CTGTGCGATGCGGCGATCATTTCCCTGCGCGCGCGCGTGTATCCCGTCAACGAGGAGG 360
DB 301 CTGTGCGATGCGGCGATCATTTCCCTGCGCGCGCGCGTGTATCCCGTCAACGAGGAGG 360
QY 361 AATTGAGTTTACAGCAAACTGATTTGACGCGCTTTGAGCGCGGATATCAGCGCTTGGGT 420
DB 361 AATTGAGTTTACAGCAAACTGATTTGACGCGCTTTGAGCGCGGATATCAGCGCTTGGGT 420
QY 421 GACTTTGACCACTGCGGATCTTGCTCAGGCGCTTCAAGATGCTATGAGGCTGCTCAA 480
DB 421 GACTTTGACCACTGCGGATCTTGCTCAGGCGCTTCAAGATGCTATGAGGCTGCTCAA 480
QY 481 CGTGAAGAGGTCAGGCTGGACTTTGAGCGGTAAGCGGTTGCTTTGAACGTTTGG 540
DB 481 CGTGAAGAGGTCAGGCTGGACTTTGAGCGGTAAGCGGTTGCTTTGAACGTTTGG 540
QY 541 GAGCGAGTCCAGAACTGATCACCATCAAGAACGCTGATTTGAGGCGCATCTATGATA 600
DB 541 GAGCGAGTCCAGAACTGATCACCATCAAGAACGCTGATTTGAGGCGCATCTATGATA 600
QY 601 TGCCACCGGCGCAACG 660
DB 601 TGCCACCGGCGCAACG 660
QY 661 CACTGCACTGAGCGCTGCTGCTGCTGGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCT 720
DB 661 CACTGCACTGAGCGCTGCTGCTGCTGGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCT 720
QY 721 GCGCGTCTACAGCAGGAGCTTTGCGCGCTGCGCAAAAGGCGCGCATCTGCTCAA 780
DB 721 GCGCGTCTACAGCAGGAGCTTTGCGCGCTGCGCAAAAGGCGCGCATCTGCTCAA 780
QY 781 CCGGCACTACTATGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
DB 781 CCGGCACTACTATGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
QY 841 ACGGATGGAATTTTCACTTGGCTGCTTGGCAATCCCATCTTTGAGAGAGGACTATCC 900
DB 841 ACGGATGGAATTTTCACTTGGCTGCTTGGCAATCCCATCTTTGAGAGAGGACTATCC 900
QY 901 AGAGAGCATGAGAGGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
DB 901 AGAGAGCATGAGAGGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
QY 961 CATCTCAATGCGGAGAGGCGCGCTTACGCGGATGAAATTAATCACTATCCGATTCG 1020
DB 961 CATCTCAATGCGGAGAGGCGCGCTTACGCGGATGAAATTAATCACTATCCGATTCG 1020
QY 1021 GCGGCACTAGCGGCTGCG 1080
DB 1021 GCGGCACTAGCGGCTGCG 1080
QY 1081 GAGAGATTAAGAGCGGCG 1140
DB 1081 GAGAGATTAAGAGCGGCG 1140
QY 1141 CCGGCACTGCTCGGAGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
DB 1141 CCGGCACTGCTCGGAGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
QY 1201 CATCAGGAGAGCGGATGCG 1260
DB 1201 CATCAGGAGAGCGGATGCG 1260
QY 1261 CAACGAGCCCTTCCGATCCGCTACTTTGACTGCGACTGAGCTCGATTTCCAAAGGCGAT 1320

DB 1261 CAACGAGCCCTTCCGATCCGCTACTTTGACTGCGACTTGAATTTCCAAAGGCGAT 1320
QY 1321 TACCAGAGCGCGCTGCTGCTCAAGGCGGATCTTGGCTGCGCGCTTGTCTGATTAATTTGA 1380
DB 1321 TACCAGAGCGCGCTGCTGCTCAAGGCGGATCTTGGCTGCGCGCTTGTCTGATTAATTTGA 1380
QY 1381 ATGGTCAATGCTACGAGACCCGATTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 ATGGTCAATGCTACGAGACCCGATTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1441 GCGGAGCCCGAAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 GCGGAGCCCGAAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 TAAAGTGGCGGCTTAAAGAGGAAATTTCTTGGCAATTTGAGCTTATGATCTTCC 1560
DB 1501 TAAAGTGGCGGCTTAAAGAGGAAATTTCTTGGCAATTTGAGCTTATGATCTTCC 1560
QY 1561 TCTCTCTTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
DB 1561 TCTCTCTTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY 1621 CCGCGGCTTCTTCTGATGAGCCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1621 CCGCGGCTTCTTCTGATGAGCCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
QY 1681 GCTCTTATGAGCTTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
DB 1681 GCTCTTATGAGCTTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1741 CTGCT 1800
DB 1741 CTGCT 1800
QY 1801 AGCAGAGTTCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1860
DB 1801 AGCAGAGTTCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1860
QY 1861 AAGGCCCATGTTTACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
DB 1861 AAGGCCCATGTTTACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
QY 1921 CCAAGGAGAGGCGCAGAGAGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
DB 1921 CCAAGGAGAGGCGCAGAGAGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
QY 1981 AAAAAAAAAA 1991
DB 1981 AAAAAAAAAA 1991
RESULT 2
US-10-026-140-3
; Sequence 3, Application US/10026140
; Publication No. US20030114330A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuurt, Frits
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BG5 Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: US/10/026,140
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-140-3

Query Match 73.0%; Score 1454.2; DB 14; Length 1455;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 ATGCCGAGTGGCTAGCTCTGCCCAAGATTGAGGGCTTCCGCAAGCGCGCTTAC 121
 DB 1 ATGCCGAGTGGCTAGCTCTGCCCAAGATTGAGGGCTTCCGCAAGCGCGCTTAC 60

QY 122 CAGATGAAAGGCGCGCTCAAGAAAGTGGCGCGCGCGCTTCCATCTGGGACAGTACTGC 181
 DB 61 CAGATGAAAGGCGCGCTCAAGAAAGTGGCGCGCGCGCTTCCATCTGGGACAGTACTGC 120

QY 182 CACCTGAGGCGATGGGCAACCAAGCGGCGCAAGCGGCGATGGCTTGGCATCTACCA 241
 DB 121 CACCTGAGGCGATGGGCAACCAAGCGGCGCAAGCGGCGATGGCTTGGCATCTACCA 180

QY 242 CGCTAGATGAGACTTTGATCTTGTACCAAGTACCGGCAAGGCTTCTCC 301
 DB 181 CGCTAGATGAGACTTTGATCTTGTACCAAGTACCGGCAAGGCTTCTCC 240

QY 302 TTGTGCTGCTGGGATCATTTCCCTCGCGGCAAGTCCGCTCAAGAGAGGA 361
 DB 241 TTGTGCTGCTGGGATCATTTCCCTCGCGGCAAGTCCGCTCAAGAGAGGA 300

QY 362 ATTGAGTTTACAGCAAACTGATTTGACGCGCTGTTGAGGGGGTATCAAGCCTGGGTG 421
 DB 301 ATTGAGTTTACAGCAAACTGATTTGAGGCGCTGTTGAGGGGGTATCAAGCCTGGGTG 360

QY 422 ACTTGTACCACTGGGATCTGCTCAGGCGCTTACGATCGCTATGAGAGCTGCTCAAC 481
 DB 361 ACTTGTACCACTGGGATCTGCTCAGGCGCTTACGATCGCTATGAGAGCTGCTCAAC 420

QY 482 GTGGAAAGGTCCTAGCTGACTTTGACGGGTATGCGAGTTGTGCTTGAACGTTTGGG 541
 DB 421 GTGGAAAGGTCCTAGCTGACTTTGACGGGTATGCGAGTTGTGCTTGAACGTTTGGG 480

QY 542 GACCGAGTCCAGAACTGGATCAACATCAAGAGCCGATTTAGGCACTTATGATAT 601
 DB 481 GACCGAGTCCAGAACTGGATCAACATCAAGAGCCGATTTAGGCACTTATGATAT 540

QY 602 GCCACCGGCAAGACGCGCGCGGCAAGAGAGCATTTAAACAAGACTTCCACGAGGCGAAC 661
 DB 541 GCCACCGGCAAGACGCGCGCGGCAAGAGAGCATTTAAACAAGACTTCCACGAGGCGAAC 600

QY 662 ACTGCCACTGAGCGCTGCTGCTGGAAGGCCCAATCATGAGCCATGCGCGCGCTG 721
 DB 601 ACTGCCACTGAGCGCTGCTGCTGGAAGGCCCAATCATGAGCCATGCGCGCGCTG 660

QY 722 GCGGCTACAGGAGGACTTTGCGCCCTGCGAAAAGGCGCAAGTCCGCACTCTGCTCAAC 781
 DB 661 GCGGCTACAGGAGGACTTTGCGCCCTGCGAAAAGGCGCAAGTCCGCACTCTGCTCAAC 720

QY 782 GCGGACTACTATGAGCCCTGGGACAGCAATGAGCTTGGGACAAAGAGGCTGCTGAGCA 841
 DB 721 GCGGACTACTATGAGCCCTGGGACAGCAATGAGCTTGGGACAAAGAGGCTGCTGAGCA 780

QY 842 CGGATGGAATTTCACTTGGCTGTTTGCATATCCATCTTTTGAAGAAGACTATCCA 901
 DB 781 CGGATGGAATTTCACTTGGCTGTTTGCATATCCATCTTTTGAAGAAGACTATCCA 840

QY 902 GAGAAGATGAAGAAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGGC 961
 DB 841 GAGAAGATGAAGAAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGGC 900

QY 962 ATCTCAATGCGGAGAGACCGACTTCTACGCGATGATTAACAACATCCAGTTGCGG 1021
 DB 901 ATCTCAATGCGGAGAGACCGACTTCTACGCGATGATTAACAACATCCAGTTGCGG 960

QY 1022 CGGCACTTAAGCGGTCCGTCGCCGAGAGCGAGCTATCTTGGCGGCACTGAGACCAAG 1081
 DB 961 CGGCACTTAAGCGGTCCGTCGCCGAGAGCGAGCTATCTTGGCGGCACTGAGACCAAG 1020

QY 1082 GAGAAATAGAGAGCGGACCCCGTTGGCGAGAGAGCGGCGCTGCTGCTGCTGCTG 1141
 DB 1021 GAGAAATAGAGAGCGGACCCCGTTGGCGAGAGAGCGGCGCTGCTGCTGCTGCTG 1080

QY 1142 CGGCACTGTTCCGGAAGCATCTGCGCGGCTGTACGCGCTGTATCGGCAAGCCATCTAC 1201
 DB 1081 CGGCACTGTTCCGGAAGCATCTGCGCGGCTGTACGCGCGCTGTATCGGCAAGCCATCTAC 1140

QY 1202 ATCAGCGAAGAGGATGCGCGCTGCGGCGTGAAGAGAAACATGACGTCGAGAGGCGCT 1261
 DB 1141 ATCAGCGAAGAGGATGCGCGCTGCGGCGTGAAGAGAAACATGACGTCGAGAGGCGCT 1200

QY 1262 AACGACCCCTTCCGATCCGCTTACTTGTGACTGCACTTGTGATTTTCAAGGCGATT 1321
 DB 1201 AACGACCCCTTCCGATCCGCTTACTTGTGACTGCACTTGTGATTTTCAAGGCGATT 1260

QY 1322 ACCGAGAGCGGCTGCTGCTCAAGGGGTACTTTGCGTGGCGTTGCTCGATACTTGGAA 1381
 DB 1261 ACCGAGAGCGGCTGCTGCTCAAGGGGTACTTTGCGTGGCGTTGCTCGATACTTGGAA 1320

QY 1382 TGSTCAGATGGCTACGAGACCCAGATTCGGCGTCAAGTCAAGACTAACACCTCAAG 1441
 DB 1321 TGSTCAGATGGCTACGAGACCCAGATTCGGCGTCAAGTCAAGACTAACACCTCAAG 1380

QY 1442 CGCAGCGCCAGAGAGTCTGCGCTGCTCTCAAGAGATGTTTGGCGCGCGAGAGGTT 1501
 DB 1381 CGCAGCGCCAGAGAGTCTGCGCTGCTCTCAAGAGATGTTTGGCGCGCGAGAGGTT 1440

QY 1502 AAGTGGCGGCAATA 1516
 DB 1441 AAGTGGCGGCAATA 1455

RESULT 3
 US-10-369-493-37092
 ; Sequence 37092, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ. ID NOS: 47374
 ; SEQ. ID NO 37092
 ; LENGTH: 1470
 ; TYPE: DNA
 ; ORGANISM: *Aspergillus nidulans*
 US-10-369-493-37092

Query Match 16.0%; Score 317.8; DB 15; Length 1470;
 Best Local Similarity 53.8%; Pred. No. 2; 6e-83;
 Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;

QY 81 TGCCCAAGCACTTGAATGGGCTTGGCAAGCGGCGCTTACGAGTCGAAGGCGCTCA 140
 DB 2 TGCCCAAGCACTTCTTCATGATATGACACGCGGCGGCGAGTTGAAGAGCGCTGCA 61

QY 141 AAGAAGTGGCCGCGGCGCTTCCATCTGGGACAGTACTGCACTTGAAGCCATCGCGCA 200
 DB 62 ACAAGATGGCAAGGATCTTGTGATCTGGGACACATTCGCGCATCTCCAGCAAGTCA 121

QY 201 CCAAGCGCGCAAGCGGATGCTGCTGCAATCACTTCAAGCGCTACGATGAGAGACTTGG 260
 DB 122 AAGACATATGCAATGGGATGACGCTGTGAGGTTTCAAGACTTTTACCGGAGAGATGTTG 181


```

QY      645 ACTCCAGGAGGACACTGCTGAGCCGTGGCTCGCTGGAGAGGCCAGATCATGA 704
      687 -----TTCCGGGGGCACTCCGCCAGCCGATCATTCGCGGACACACTCCTCTCG 741
QY      705 GCCATGCCCGCGCCGTGGCCGTCTACAGCAGGAGCTTCCGCTCCCAAAAGGCCAGA 764
      742 CGACCGCTGCGCTGCTCAGCTGACCGCAGCAAGTCAAGATTGTCAGCGCCGCAAGA 801
QY      765 TCGGCACTCTGCTCAACGCGGCACTACTATAGAGCCCTGGAGCAGCACTTCGAGCA 824
      802 TCGGCACTCTGCTGAGCTTGTGTGTGATGAGCC--GCTAGCAAGCTCGGTGAGGACG 858
QY      825 AGGAGCTGCTGAGCGACGAGTGAATTTCATTTGGCTGTTTCCCAATCCATCTTCT 884
      859 AGTACGGGGGCGACCGGGGCTCGAGTGTTCACCTTGGCTGTTCTTCACCCCATCACT 918
QY      885 TGAAGAAGACTATCCAGAGAGCATGAAGAAGCAGCTGGCGAGAGGCTTCCAGCTTCA 944
      919 ---ACGGCCACTACCCCGGAGACGATGCAAGATCGTCATGGGGAGGCTGCCCAACTTCA 975
QY      945 CTCGGCGAGCTTGGCATCTGCTCAATGCGGAGAGACGCACTTCTAGCGATGAATTACT 1004
      976 CTTTGAAGCACTGCTGAGCTTGTGTGTGATGAGCC--GACTAGTTCCTCATACCACT 1032
QY      1005 ACACATCCCACTTCCGCGCGCCACCTAGACGAGTCCGCTCCCGAGACGCACTATCTCGCG 1064
      1033 ACACAGCACTACTAGCGCAGCACTTGTCAACGCCACAGAGACCACTACCGCAAGATT 1092
QY      1065 CCATCCATGACACACAGAGAAATBAGACGCGACGCCCTGTTGGCGAGAGAGCGGCTCG 1124
      1093 GGAATGCAAGATTTCTATGAGCGAGATGTGTGCCATTTGGCAAAAGGCGTACTCGG 1152
QY      1125 CTTGGCTGGGCTCTGCGCGGAGACATGTTCCGGAAGACATCTCGCCGGGTGTACGGCTGT 1184
      1153 ACTGGCTTTAGCTGTCTTCAATGGGGGCTTACAAAGGCTCTGATTTGAGCAAGAGAGAT 1212
QY      1185 ACAGGCAAGCC---CATCTACATCACCGAAGCGAGATCCCGTCCCTGAGAGAGAGACA 1241
      1213 TCAACAGCCCTGTGATGCTCATCGAGAGAAACGGAATTGACCAAGCTCGAAGAAAGACCT 1272
QY      1242 TGACGTGCGAGAGGCGGCTCAAGACCCCTTCGCGATCCGTAACCTTGACTCGCACTTG 1301
      1273 TGCCTGTTCC-----GCTCTGTAGCAGCAAGTTTCAAGATAGACTACTTGCAGAAAGTACTGT 1326
QY      1302 ACTCGATTTTCAGAGCGCATTAACCCAGAGCGGCGTGTCTCAAGGGTACTTTGGCGGTG 1361
      1327 ACAGGCT---CCAGTGGCCATACGCGACGCTGCAAAAGTCTTCCGCTACTTCCGCTGT 1383
QY      1362 CGTTGCTCGATTACTTGAATGGTACAGATGGCTACCGAGCCAGATTTCGCGCTCACTTCA 1421
      1384 CGCTGCTGAGCAACTTGCAGATGGCGGCTCGGCTTCACTCCAAAGTTTGGATCGTGTACG 1443
QY      1422 CAGACTACACCACTTCAGCGGACGCGCCAAAGATGTGCTGCTGCTCAAGGA 1476
      1444 TGAGCGGAAACAGTGTGCTCGGATACCCCAAGACTCAGCTCGCTGTTCAAGAA 1498

```

```

RESULT 5
US-10-425-114-19747
; Sequence 19747, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven B.
; APPLICANT: Tabaek, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28

```

```

; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 19747
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB151-012-C9_FLI
; US-10-425-114-19747

Query Match      11.6%; Score 230.4; DB 12; Length 1794;
Best Local Similarity 52.3%; Pred. No. 2,4e-57;
Matches 722; Conservative 0; Mismatches 616; Indels 42; Gaps 8;

QY      101 GGCTTGCAACGCGCCCTTCTCCTTGTGCTGCGGATCATTTCCCTCGGCGGACAGCTG 160
      237 GGCACCGCCACTCTCTGTCAGATGCAAGAGGCGCTACTCTGAGGCAAGAGCTTGG 236
QY      161 TCCATCTGGGACAGTACTGCACTGAGCCATCGCGACCAACGCGCCCAACGCGAT 220
      297 AGCAACTGGGATGTCTTTCAGCCACGTACAGAAAGATCGAAGACGGAACCGGAGAT 356
QY      221 GTGGCTTGGATGATCACTACCAACCGCTACGATGAGACTTGTATCTTTCACCAAGTACGGC 280
      357 ATGCGGATGATCATATCATATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 416
QY      281 GCAAGGCTTACCGCTTCTCTTGTGCTGCGGATCATTTCCCTCGGCGGACAGCTG 340
      417 ACCAAGCTTACAGATTTTCTATATATATGAGCCAGAAATTTCTTC---GAGAGGAGATTTC 473
QY      341 GATCCGCTCAACGAGAGGAGAAATGAGTTTACAGCAAACTGATGACGCTGTTGAGG 400
      474 GCGAGAGTCAACCAAGAGCATCGGCTTCTACACAGACTCATGCTGCTCTGCTC 533
QY      401 CGGGATATACGCTTGGGATTTGTATCATATGAGATGTGCTTACAGGCTTTCAGAT 460
      534 AAGGATAGAGCGCTTGTGAGCTGTCCACTACAGCAATCCCGAAGAGCTGAGAGAC 593
QY      461 CGCTATGAGAGCTGCTCAACGTTGAGAGAGTCCAGCTGAGACTTTGAGCGGATG 520
      594 AGGTACGGGCGTGTGCTCGGCGGCGCGGAGCGCGGCGGACTTTCGCGACCTGCGGAC 653
QY      521 TTGTGCTTTGAAAGTTTGGGAGCCGAGTCCAGATGATGATCAACATCAAGACCTG 580
      654 GTGTGCTTGGCGGCTTGGCGGACCGGTCAGCACTGGGCGCACTTCAACGAGCCAAC 713
QY      581 ATTACAGGCTATATGATATGCAACCGGACGAGACGCGCGGCGAGAGACAGATTAC 640
      714 GTGCGGTCAACGAGGAGCTACATGCTCGGACCTACCGCGGCGGAGATGC-----TCG 767
QY      641 AAGCACTCAACGAGGAGCAACATGCACTGAGCGGTGCTGCTGGAAGAGCCAGATC 700
      768 CGCTCTGCGCCCGGAGCACTCGAGCGGAGCCCTTACGTCCCGCGCAACAGTGTTC 827
QY      701 ATGAGCCATGCGCGCGCGCTGCTCTACAGCAGGAGACTTTCGCCCTTCGCAAAAGGC 760
      828 CTGGCCAGCGCGCGCGCTGCAAGTCTAAGAA-----CAAAGTACAGAAAGGC 878
QY      761 CAGATGCGATCTGCTCAACGCGGACTATATGAGCCCTTGGACAGCAATGAGCTCG 820
      879 TTGATGCGGATCTGATGATCTCACTGTCTGTTGTGCTCGCTG---ACGAGCGCGCGCG 935
QY      821 GACAAGAGGCTCTAGAGGAGGAGATTTCAATTGAGCTGCTGCTGCTGCTGCTGCTGCT 880
      936 GACCGGCTGCGCTAGAGGAGGAGGCTGCTTGAAGTTCATGATTTCTTGAACCGGATA 995
QY      881 TTCTTGAAGAGACTATTCAGAGACATGAGAGACAGCTGAGCGAGAGCTTTCAGCC 940
      996 ATCT---AAGGAGACTACCTTCGAGAGATGCGCGGCTCTGAGGCTTCAGAGCGGAGC 1052
QY      941 CTACTCTCCGCGGACTTTGCAATTCCTCAATGCGGAGAGACGAGACTTTCAGGCAATGAAT 1000
      1053 TTCTTCCCGGAGATGAGAGAGCT---GGGCTAGGAGCTGAGACTTTCATCGGCACTAAC 1109

```

```

QY 1001 TACTACATCCAGTTGCG-----GCGCACCTAGACGCTCCCGTCCCGAG 1048
DB 1110 CACTACACGACGCTGTAACGACGAGATGTGATTTCTCGCAGATTATGCCGTCGGGG 1169
QY 1049 ACAGACTATCTCGGCGCCATTCATGACACACGAGAGATAGACGCGACCCCGTTGGC 1108
DB 1170 CAGAGATTCCACCAATCGCTGGCGCCCTACACGAGAGAGAGACGCGATTCCATCGGG 1229
QY 1109 GAGAGAGCGGCCCTCGCTGCGCTCCGCGCCGACATGTTCCGGAAGCATCTCGCC 1168
DB 1230 CTTCCGACAGAGATGCGACCTTTTACGTGTTCTGACCGGATAGAGAGATGCTACCC 1289
QY 1169 CGGCTGTAACGCGCTGTGACGCA---GCCCATTAATCAACGAGAAAGCATCCCGTC 1225
DB 1290 TACATCATGAAACAGATACAGCAATCTCCCATGTTCAATACGAAATGCTACGCAAA 1349
QY 1226 CTTGAGAGAGAAACATGACGTGCGAGAGAGCCGCTCAACGACCCCTTCGATCCGATAC 1285
DB 1350 GGTGGGAGTGGTTTATCTAACCCGTGCGAGGACGTGGCTTGACACGAGGACGATACAG 1409
QY 1286 TTTGACTGCACTTGGACTCGATTTCAGAGCCATTAACGAGACGCGCTGTCTGCAAG 1345
DB 1410 TACCTGAGGGCTACCTGCAAACTGCGCAAAAGTCATGAGTACGCGCGCGATGTCGCT 1469
QY 1346 GGGTACTTTGCGTGGCGCTTGTGATGATCTTGGAATGCTGATGCTAGGACCGAGA 1405
DB 1470 GGCCTACTTCACTGTGCTCTCTCATGACAAATTCGAGTGTATGAGGTACACCTCTCGC 1529
QY 1406 TTGCGGCTCACTGTCACAGACTACACGACCCCTCAAGCCGACGCCCAAGAGTCTGCGCTG 1465
DB 1530 TTGCGCTTCACTACGTTGATCTACCAACGACGAGAGAGAACCCCAATCGTGGCGCTG 1589

```

RESULT 6 US-10-425-114-30939

```

; Sequence 30939, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 73128
; SEQ ID NO 30939
; LENGTH: 1781
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73106C01_FLI
US-10-425-114-30939

```

Query Match 11.4%; Score 227.4; DB 12; Length 1781;
Best Local Similarity 52.2%; Pred. No. 1,9e-56;
Matches 735; Conservative 1; Mismatches 637; Indels 36; Gaps 9;

```

QY 81 TGCCCAACGACTTGAATGGGGCTTGCACACGCGCCCTACCAAGATGAGAGCGCGCTCA 140
DB 230 TCCCAAGGGGTTGTGTTGCGAGCGCGACGTGCGGTACAGGTGAGAGGGCGCGCT 289
QY 141 AAGAAGTGGCGCGCGCGCGCTGCTGATGAGACATGATGCTGACCTGAGACCATCGCGCA 200
DB 290 CCACCAACGCGCGCGCGCGCTGCTGATGAGATGATGCTGACCTGAGACCATCGCGCA 349
QY 201 CCAACGCGCGCAACGCGCGATGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 260
DB 350 CAGGAATCAAAATGAGACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 409

```

```

QY 261 ATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTGTCGTGTCGGGATCA 320
DB 410 ATCTTATATAAAGTTTGAATCTTTGATGCTTACCGCTTCTCAATCTCATGATGATGATGATGAT 469
QY 321 TTCCCTTCGCGGAGCGGTGATCCGTCACAGAGAGAGGATGATGATGATGATGATGATGATGAT 380
DB 470 TCCCGATGAGGAGAGG-----AAGTCATCCAGAGAGGATGATGATGATGATGATGATGATGAT 523
QY 381 TGAATGACGCGCTGTTGAGGCGGGTATCAAGCGCTTGGGTGATCTTTATACCTGAGATC 440
DB 524 TGATTAATATCTTGTTCAGAGAGCATGATCTCTTATCATTAACCTTTACCATATGATTC 583
QY 441 TGCCTCAAGCGCTTACCATGCTATGAGAGCTGCTCAACGTGAGAGAGGTCCAGACTGG 500
DB 584 TTCTCTTTCGCTTGAAGAGAAATATGAGGATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAG 642
QY 501 ACTTTGAGCGGTATGAGAGGTTGCTTGAACGTTTGGGAGCCGAGTCCAGATCTGCA 560
DB 643 --TTTACAGACTATGCTGACTTCTGTTTAAGACCTACGCGGATGCTTAAAGACTGGT 700
QY 561 TCACCATCAAGAGACCTGATTCAGGCACTATGATATGATGATGATGATGATGATGATGATGAT 620
DB 701 TTACATTCATAGGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 760
QY 621 CCGGACAGAGCAATTAACAGCACTCCACGAGGCAACCTGCTGACGCGTGGC 680
DB 761 CTCAAAGGTG-----ACCATGCGCTGCTGAGGAGATTCAGCAACCGAAGCTTACA 814
QY 681 TCGCTGAGAGAGCGCCAGATCATGAGCATGCGCGCGCTGCGCTTACAGAGAGACT 740
DB 815 TATGTTCTATATATTTTCTTGTGACATGCTATGCTGATGATGATGATGATGATGATGATGAT 874
QY 741 TTGCGCTTGCAGAAAGGCGCAAGATGCGATCTGCTCAACGCGCATCTATGAGCCT 800
DB 875 ATGAGCGTGTCTAGAGAGGATGAGTGAATGCTGCTGATCTTCACTGATGAGAGCTTC 934
QY 801 GGAACAGCAATGAGCTTGGGACAAAGAGCTGCTGAGGACGAGATGATGATGATGATGATGAT 860
DB 935 T--TACAACTGACCTGATGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 991
QY 861 GCTGTTTGCATATCCCATCTTCTTGAAGAGAGACTATGAGAGAGAGAGAGAGAGAGAGAGAG 920
DB 992 GCTGTTTGTGATGCT--ATTGATAAGGAGACATATCAACAGATGATGATGATGATGATGAT 1048
QY 921 TGGCGAGAGAGCTTTCAGCCCTCACTCCGCGAGCTTGGCATCTTCAATGCGGAGAGA 980
DB 1049 TGAAGAGAGAGCTGCGCAGGTTCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1108
QY 981 CCGACTTCTACGAGATGATTAAC--TACATATCCAGATTGCGCGCACTTGAAGAGGTC 1037
DB 1109 ACTACATGCTGATTAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1168
QY 1038 CCGTCCCGAGAGAGAGATCTGCGCGCATTCATGACACAGAGAGATTAAGAGAGCGCA 1097
DB 1169 TGGGCGCAAGTATGCTGCTGCGAGTGGAGGTTCAATATGTTTTCAGCGCAATGCA 1228
QY 1098 GCGCGCTTGGGAGAGAGAGCGCGCTCGCTGCGCTGCGCTGCGCGAGATGTTCCGA 1157
DB 1229 AACCATTTGACCAACAGGAGATTTAAATGAGCTTCAATGCGCGCGAGATGATGAG 1288
QY 1158 AGCATCTGCGCGGATGATGAGCTGATGAGAGAG--CCATTTACATCAACGAGAGAG 1214
DB 1289 GGTGCTGATATCTACTTCAAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAG 1348
QY 1215 GATGCGCTGCTTGGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1274
DB 1349 GATGAGACAGAGCTGAG-----AATGAGCCGAGAGCAATTAACCTGCGAGAGCGCA 1402
QY 1275 GATATCGGATCTTATCTGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1334
DB 1403 GGTGCGGTTTACAGAGAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1459

```

QY 1335 TCGTCGCAAGGGGACTTTCGTCGGCGCTTGCATTAAGTTGAAATGTCAGATGGCT 1394
DB 1460 CGAAGCTGGCTGCTACTTCCGCTGCTCTCTCCGCAACTTCGAGTGGCTGGCAGGGT 1519
QY 1395 ACGGACCCGAGTTTGGGGGTCAAGTTCACACTACACCCCTCAAGCGACGCCAAGA 1454
DB 1520 ACTGCTCAAGTTGCGGATGCTCTACGTGACCTTCAACAGCTGAAACGCCACCCGAAG 1579
QY 1455 AGTCGCCCTGCTCTCAAGGACATGTTT 1483
DB 1580 CGTCGGCTTACTGTTTCAAGGACATGCTT 1608

RESULT 7
US-10-425-114-30835
; Sequence 30835, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaka, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30835
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: UC-ZMFLB73086G03_FLI
US-10-425-114-30835

Query Match 11.4%; Score 227.4; DB 12; Length 1931;
Best Local Similarity 52.2%; Pred. No. 2e-56;
Matches 735; Conservative 1; Mismatches 637; Indels 36; Gaps 9;

QY 81 TGGCCACGACCTTGAATGGGGCTTCCCAACGGCGGCTCAACAGATCGAAGCGCGCTCA 140
DB 236 TCCCAAGGGGTTGTTGGGACGGGCAAGTGGGTAACAGTGGAGGCGCGCGCT 295
QY 141 AAGAAGTGGCGCGCGCTCCATCTGGGACAGTACGCCACCTGAAGCATCGCGCA 200
DB 296 CCAACCAAGGCGCGCGCTCCATCTGGGATTCATTGCGGCAAGTCCAGGAATATTG 355
QY 201 CCAAGCGCGCAACGGGATGGGCTTGCATCTCAACCGCTAGATGAGAGACTTNG 260
DB 356 CAGGGAATCAAAATGAGAGGTTGCAAGTGAATCAATCGCTACAAAGAGACGCTG 415
QY 261 ATCTTTGACCAAGTACGAGCGCAAGGCTTACCGCTTCTCTGTCGTCGCGATCA 320
DB 416 ATCTCAAGAAAGTTGAATTTGATGCTTACCGGCTTCTCATCTATGATCCAGATCT 475
QY 321 TTCCCTCGCGCGGAGGCTGATCCCTCAACGAGGAGGAATTAAGTTTAAACGAAAC 380
DB 476 TCCCGAGTGGCGAAGG-----AAAGTCAATCCAGAGGTGTAGCGTATTAACAATATT 529
QY 381 TGAATGACGCGCTTGAAGCGGGGTATCAAGCTTGGGATGCTTTGATCACTGGATC 440
DB 530 TGAATAACTATCTGCTTCAGAGGCAATGACTCTTACATCACTTTTACCATATATC 589
QY 441 TGCCTCAGCGCTTCAAGATCGATGAGAGGCTGCTCAACGTAAGAGAGTCCAGCTGG 500
DB 590 TTCTCTTGGGCTTGAAGAAATATAGAGGTTGTTAAGCGGCAATATGGCGGACTTG- 648
QY 501 ACTTTGACGGTATGCAAGGTTGCTTTGAACCTTTTGGGAGCGAGTCCAGAACTGGA 560
DB 649 -TTTACAGACTATGCTGACTTCTGTTTAAAGCACTACGGGATCGGTAAGCACTGCT 706

QY 561 TCACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGGACGACGCC 620
DB 707 TTACATTGATGAGCCAGGATAGTAGCCGCTACTTGGCTATGACAGGGTCAAAATCTC 766
QY 621 CGGCGAGGAGCAGCAATTAACAGCACTCCACGAGGCAACTGCCACTGAGCGCTGGC 680
DB 767 CTCAAAGGTG-----ACCAGATGCGCTGCTGTTGGAAATTAAGCAACGAACTTACA 820
QY 681 TCGCTGAAAGGCCAGATCATAGAGCATGCCGCGCGGCGCTGTACAGCAGGACT 740
DB 821 TAGTGTCTATTAATTTCTCTTGGCAATGCTACTGAGTTGAAAGTATCCGTAAGAAAT 880
QY 741 TTGCCCCCTGCAAAAGGCGCAGATGCGCATCTGCTCAACGGCGACTACTATAGCCCT 800
DB 881 ATCAGGTGCTCAGAGAGGTTAAGTGGATAGTATGCTCGTCACTTAACTGTTACAGAGCTC 940
QY 801 GGGACAGCAATGAGCTCGGGAACAAGAGGCTGTGAGCGACGAGTGAATTTCACTTG 860
DB 941 T--TACAACTCAGCTGATGACCAAGCAGCAGCCCAAGAGCCAGGACTTCCACTTG 997
QY 861 GCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCAGAGAGATGAAGAGAGC 920
DB 998 GCTGTTTGTGATCC--ATTGATTAACGAGACATATCCAGATTAATGCAAGATCTCG 1054
QY 921 TGGGCGAGGCTTCCAGCCCTCACTCCCGGACTTGGCATCTTCAATGCGGAGAGA 980
DB 1055 TGAAGAGAGGCTGCCAGGTTCACTCTGACAGGCTAACTGTGTAAAGGCTCGGCAG 1114
QY 981 CCGACTTCTACGGCATGAATTAC--TACACATCCAGTTCGCGCGCACCTTACAGCGTTC 1037
DB 1115 ACTACATCGGTATCAACGAGTATACATCCAGTACATGAAGGGGCGAAGACTGGTCCAGC 1174
QY 1038 CCGTCCCGGAGAGCACTATCTGCGGCCCATTCATAGCAACGAGAAATGAAGACGCA 1097
DB 1175 TGGGCGCAATCTCTGCGATTTGGCAGGTTCAATATGTTTTTTCACGCAATGCA 1234
QY 1098 GCCCGGTCGAGAGAGGCGGCTGCGCTGCGCTCTGCGCGGACATGTTCCGGA 1157
DB 1235 AACCGATGGAACACAGGCGAATTTCAATGCTCTTACATCGCCGACGGGATGTAAG 1294
QY 1158 AGCATCTGCGCGGGGTATACGGCTGTACCGCAAGC--CATCTACATCACGAGAACG 1214
DB 1295 GGTGCGTGAATCTCTCAAGAGAGATGATGGAATCCAAAGATCTACATTAACGAGAACG 1354
QY 1215 GATGCCGCTGCTGAGAGAGAACTAGCTGCGAGAGGCGCTCAACGACCCCTTCC 1274
DB 1355 GAATGACCAAGCTTGA-----AACTTACCCGAGACCAAGTACTGCGGAGCGCACGA 1408
QY 1275 GCATCCGTAATTTGAATGCACTGCACTTGAATTCGAAGGCATTAACCGAGAGCGCG 1334
DB 1409 GGGTGGGTTCTACAGAGATCATGCGCAGCTGAAGAGGCAATAGACAG--GGAG 1465
QY 1335 TCGTCTCAAGGGGTACTTTGCGTGGCGTGTGCTGATTAATTTGAATGTCAGATGCT 1394
DB 1466 CGAAGCTGCTGCTACTTCTGCTGATCTCTCTGCAACTTTCGATGCTGGAGGCT 1525
QY 1395 ACGGACCCGAGATTCGGGCTGACGTTACAGAGCTACACACCTTCAAGCGGACGCCAAGA 1454
DB 1526 ACTGCTCAAGTTGCGGATGCTGTACGTGACCTTCAACAGCTTCGAACGCCACCGAAG 1585
QY 1455 AGTCGCCCTGCTCTCAAGGACATGTTT 1483
DB 1586 CGTCGGCTTACTGTTTCAAGGACATGCTT 1614

RESULT 8
US-10-425-114-33175
; Sequence 33175, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua

```

1  APPLICANT: Kovalic, David K.
2  APPLICANT: Screen, Steven E
3  APPLICANT: Tabaska, Jack E
4  APPLICANT: Cao, Yongwei
5  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
6  TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
7  FILE REFERENCE: 38-21 (5313) B
8  CURRENT APPLICATION NUMBER: US/10/425,114
9  CURRENT FILING DATE: 2003-04-28
10 NUMBER OF SEQ. ID NOS: 73128
11 SEQ. ID NO 33175
12 LENGTH: 1932
13 TYPE: DNA
14 ORGANISM: Zea mays
15 FEATURE:
16 OTHER INFORMATION: Clone ID: UC-ZMFLMOL7071E02_FLI
17 US-10-425-114-33175

```

Query Match	11.4%	Score 227.4	DB 12	Length 1932
Best Local Similarity	52.2%	Pred. No. 26-55		
Matches	735	Conservative	1	Mismatches 637
			Indels	36
			Gaps	9
QY	81	TCGCCAAGCACTTGAATATGAGGCTTCGGAAACGGCCGCTTACAGATCGAAGGCGCCGTCA	140	
DB	237	TCCCAAGGGGTTCGTGTTCCGGACGGGACGTCGGGATACAGATCGAAGGCGCCGCT	296	
QY	141	AAGAAGTGGCCGCGGCGCCGTCATCTGGGACACGTACTGCCACTCGAAGCATGCGCA	200	
DB	297	CCACCAACGGCCGGGGCCCTTCATCTGGGATTCATTGCGGACATCCAGGAATATTG	356	
QY	201	CCAAAGGGGCGCAACGGCGATGNGGCTTGGGATCATACCAACGCTACGATGAGGACTTTG	260	
DB	357	CAGGGATCAAAATGAGACGTTGACGTGGATCAATACATGCTGCTACAGAAAGACGTG	416	
QY	261	ATCTCTTGAACCAATGACGGCGAAAGCCCTACCGCTTCTCTGTGCTGTCGGATCA	320	
DB	417	ATCTCATGAAAGTTTGAATTTGATGCTTACCGGTTCTCAATCATGATGTCAGATCT	476	
QY	321	TTCCCTCGGCGGACGCTGATCCCGTCAACGAGAGGGAATTAGTTTACAGAAAC	380	
DB	477	TCCCGGATGGCAAGGG-----AAGTCAATCCAAAGGTGAGGTATTCATATATT	530	
QY	381	TGATTTACGCCCTGTGTGAGCGGGGATATACGCTTGAGCTTTGTAACCATGGGATC	440	
DB	531	TGATTAACATATCTGCTTCAGCAAGGATGATCTTACATCAACCTTTACATATGATC	590	
QY	441	TGCTCAGGCGCTTACGATTCGCTATGAGGCTGCTCAACGTGGAAGGTCACGCTGG	500	
DB	591	TTCCCTCTTGCGCTTGAGAAATATGAGAGGCTGTTAAGCGGAGATGGCGGACTTG	649	
QY	501	ACTTTAGCGGATATGGAGGTTGCTGTTTGAACGTTTGGGACCGAGTCCAGAACTGGA	560	
DB	650	--TTTACAGACTATGCACTTCTGTTTAAAGCTTACGGCGATCCGTTAAAGCATGGT	707	
QY	561	TCACCATCAACGAGMCCCTGATTCAGGCGCATCTATGATATGCAACCGGACGAAGCCC	620	
DB	708	TTCATTTCAATGAGCCAAAGATATGTAAGCGTACTTGGCTATGACACAGGCTCAATCTC	767	
QY	621	CGGCGAGGACGACATTAACAGCACTCCACGAGGGCAACACTGCCCTGAGCCGTGGC	680	
DB	768	CTCAAAAGTGC-----ACCAAGATGGCGTGGCTGATGGGAATTAGCAACCGAACCTTCA	821	
QY	681	TGCTTGGAAGGCCAGATCATAGAGCATGCCGCGCGCGTGGCGCTACAGCAGGAGCT	740	
DB	822	TATTTGCTCAATATTTTCTTTGGCAATCTATGCACTTGCAGTTGCAAGTACCGTACGAAT	881	
QY	741	TTGGCCCTTCGCAAAAGGCGCAATTCGGACTTCGCTCAACGGGCACTACTATGAGCCT	800	
DB	882	ATCAGGTGCTCAGAAAGGATTAAGGTGGAAATAGTCTGACTTCAATGTGTAAGAGGCTC	941	
QY	801	GGAGCAGCAATGAGCTTCGGGACACAGAGCTGTGAGCGACGATGGAAATTTTCAACTG	860	
DB	942	T---TTCAACTCACTGATGACCCAAACACAGCCCAAGAGCCAGGAACTTCCACTTGG	998	

QY	861	GCTGGTTTGC	CAATCCCATCTTCTTGA	GAAGGCTATCCAGAGAGCATGAAGAGC	920	
Db	999	GCTGGTTTGT	TGATTC---ATTGATTA	AACGACATATCAAGATATGACATCTCG	1055	
QY	921	TGGGCGAGAGGCTTCCAG	CCCTCACTCCGCGAGCTTTCGCATCTCAATGCCGAGAGA	980		
Db	1056	TGAAGAGAGGCTCCAG	GTTCACTCTCTAGCAGGCTTAACCTGTGA	GGCGCTGGGAG	1115	
QY	981	CCGACTTCTA	GGGATGAATTAC---TACA	CATCCAGTTGCGCGGCACCTTAA	CGGTC	1037
Db	1116	ACTACATCGATCA	AGAGTACATTCACAGCTCATGA	AGGGGACAGAGCTGTTCAGC	1175	
QY	1038	CCGTCCTCCGAGACGGA	CTATCTGGGCGCATTCATGA	GACACAGAGAATTAAGACGCA	1097	
Db	1176	TGGGCGCCATGAGTACTCTG	CCCATTTGGCAGGTTCAATATGTTTTC	ACGCAATGGCA	1235	
QY	1098	GCCCCGTGTGCGAGAGAGCG	CCCTCGCTCGCTCTGCCCCGACATGTTCCGGA	1157		
Db	1236	AACGATTTGAC	CAAGCGAATTTCAATGAGCTCTACATCGCCCGACGGAGTACG	1299		
QY	1158	AGCATCTCGCCCGGCTG	TACGCGCTGTACGGCAAGC---CCATCTACATCAACGAGAG	1214		
Db	1296	GGTGCGTAACTACCTCA	AGAGAGATGTGGAAATCCAA	GCATCTACATTAACGGAACG	1355	
QY	1215	GATGCCGCGCCCTGAGAGAGAA	CATGAGTGCAGAGAGGCGTCAACGACCCCTTCC	1274		
Db	1356	GAATGGACCAAGCTGA	-----AACTTGACCGAGACCAATACCTCGGCACGCA	1409		
QY	1275	GCATCCGTA	CTTTGACTCGACTTGGACTCGATTTCCAA	GGCCATTACCGAGACGGCG	1334	
Db	1410	GGGTGCGGTTCTACAGAGACT	ACATCGCCAGCTGAAAGGCCCATAGACAG---GGAG	1466		
QY	1335	TGCGTGTCAAGGGGATCTTGG	CGGGGCGTGTGCTGATTAATCTGGAAATGTCA	TGAGT	1394	
Db	1467	CGAAGTGGCTGGCTACTT	CGCTGTGTCTCTCTGCA	CAATTCGAGTGGCGGAGGCT	1522	
QY	1395	ACGAGCCGAGATTTGGCG	GTCAAGTTCAAGACTACACACCTCAAGGCA	CGCCCAAGA	145	
Db	1527	ACTCGTCAAGTTGGC	ATCGCTACGTGAGACTTCACACAGCTCGAA	CGCCACCCGAAG	1586	
QY	1455	AGTTCGCTCGGTCTCA	AGGACATGTT	1483		
Db	1587	CGTGGCTTACTGTTCA	GGGACATGCTT	1615		

```

RESULT 9
US-10-425-114-25892
; Sequence 25892, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 25892
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB960-023-C7_FLI
US-10-425-114-25892

Query Match      11.0%; Score 219.2; DB 12; Length 1737;
Best Local Similarity 51.3%; Pred. No. 5e-54;

```

Matches 727; Conservative 1; Mismatches 654; Indels 36; Gaps 8;			
Qy	81	TGCCAAGCAGCTTGAATGGGGCTTGCACAGCGCGCTTACGATCGAAGGGCGCTCA	140
Db	228	TTCCCAAGGGGCTTGTGTTGGGACGGCGGCTTGGCTTACGATGGAAGGACATGGCGC	287
Qy	141	AAGAAGTGGCCGCGGCGCTTCCATCTGGGACGTAATCCCACTGGAGCCATCGGCA	200
Db	288	ACAAGAGCGCGCGGCGCGAGCATTTGGGACGCTTATCATAGATCCCGGGGAGATCG	347
Qy	201	CCAAAGCGCCCAAGCGGATGTGGCTTGCATCTACACCGCTACGATGAGACTTTG	260
Db	348	CAAAACAACCGCACCGGAGAGTGTGACGATCTGTCATCTGCTACAAAGAGACCTGA	407
Qy	261	ATCTCTTGAACCAAGTACGGCGCAAGGCTTACCGCTTCTCTGTGTGTGGTCCGATCA	320
Db	408	ATATCTATGAAGATATGGGCTTGCAGCGCTTACCGGTTTGTGATCTTGTGTCAAGATAT	467
Qy	321	TTCCCTCGCGGCGAGGCTGGATCCGCTCAACGAGAGGAGATTTGATTTTACAGCAAC	380
Db	468	TCCCAATGGAAAC-----GGAAGTGAACGGAAGAGAGTGGGTACTACAAACGCG	521
Qy	381	TGATTTAGCGCTTGTGAGCGGGGATTCAGCGCTTGGGTGACTTTGATCCACTGGATC	440
Db	522	TGATTAACATCATGTGTGAAGAAAGGATCAACCTTACGCAACCTGTACCTACGACC	581
Qy	441	TGCTCAGGCGCTTACGATCGCTATGAGAGCTGGCTCAACGTAAGAGATCCAGCTGG	500
Db	582	TACCGAGGGCTGAGAGCTCGGTACGAGAGGCTGTGACAGAGAGTGT--GAGAT	638
Qy	501	ACTTTAGCGGATATGCGAGGTTGTGCTTTGAAAGTTTTGGGAGCCGAGTCCAGAACTGA	560
Db	639	CGTTGCGACATACGCGCACTTGTGCTGGGGCGTTCCGCGACAGGATGAAGACTGGC	698
Qy	561	TCAACATCAACGAGCCCTGGATTCAGGCCATCTATGATATGCAACCGGACGCAACGCC	620
Db	699	TGACGTTCAACGAGCGCGGGTGTGTGCTCCCTAGGGTTCGACGACGCGAGCTTGGCGC	758
Qy	621	CGGGGAGGAGACATTTAACAACACTCCACGAGGCGCACTGCACTGAGACCTGGCG	680
Db	759	CGGGAGGGTGCACGGGGTGCAG--GCGGGGGGAGCTGGGACACGAGCCCTTACG	812
Qy	681	TGCTGGAAGGCGCCAGATCATGAGCCATGCGCGCGCTGGGCGCTTACAGAGGACT	740
Db	813	TGCTGGCGACCACTCATCTCTCCACGCGCGCGCTGCGAGAGTACCGCGCAGGC	872
Qy	741	TTGCGCCCTTGCAGAAAGGCGCAGATCGGATCTGCTCAACGGCGACTATGAGCCCT	800
Db	873	ACCAGCGCAGCGAGAGGGGCGAGGCTGGATCTGCTGAGATTTGATGTGTACGAGCCC	932
Qy	801	GGGACAGCAATGAGCTCGGAGCAAGAGGCTGTGAGCGACGATGGAATTTCACTTG	860
Db	933	TCAAGCGGAGCTCAGCGCGCGAGCGCGCGCTCAMAAGTCCAGAGATTTCAAGTGC	992
Qy	861	GCTGTTTCCATCCCATCTTCTTGAAGAAGACTATCCAGAGAGCATMAAGACAGC	920
Db	993	GATGTCTCTGACACCCATCTCT---ACGGCGAGTACCCCAAGTGGGTCCGAGAACG	1049
Qy	921	TGGGCGAGAGCTTCAAGCCCTCACTCCGCGGACTTTGCGCATCTCAATGCGGAGAG	980
Db	1050	TCAAGGCGAGGCTCCCAAGTTCAAGCTGTAGGA---GGCGGCTTATGTCGAGGCTCCA	1106
Qy	981	CGGACTTCAACGCGCATGAAATTAACAATCCAGTTTGGGCGCCACTAGAGAGGCTCCG	1040
Db	1107	TGCACTACGTCGAGTCAACAGTACATGCTTACTACGTCGTGTATGACGCGCAAG	1166
Qy	1041	TTCCCGAGAGC-----GACTATCTCGGCGCCATCTCATGACACAGAGAAATAGG	1091
Db	1167	CTACGGCGGGCGCGCCAGCTACTCGTCCGACTGGCAGCGTGTGTGTATGAACGCG	1226
Qy	1092	ACGGCAGCCCGTTGGCGAGAGAGCGGCTGCTGGCTGCGCTCTGCGCGAGCATGT	1151
Db	1227	ACGGTGTGCGATTGAGCAAGGCGCAACTCAAGTGGCTCTACATGTGCTTGGGAGC	1286

Qy	1152	TCCGAAGCATCTCGCCGGGTGTACGCGCTGTAGCGCAAGCCCATCTACATCAGGAGA	1211
Db	1287	TGTACAAAGCCGTACCTACTACATGAGAGAGTACGCAACCCCA---CGATCTCTGT	1343
Qy	1212	ACGATGCCCGTGGCCCTTGAGAGAGAACATGACGTGCGAGAGGCGCTCAACGACCTT	1271
Db	1344	CAGAGAACGGTATGACGACCCCGGCAAGTCAAGTGGGCGAGGCGGTGCGGACGCGG	1403
Qy	1272	TCCGATTCGGTACTTTGACTGCACTTGGACTGCAATTTCCAAAGGCAATTAACAGAGC	1331
Db	1404	CGAGGTGCGCTACT---ACCGAGCTACGTGCGCAGCTGAAGGCGCGCATGACGCGC	1460
Qy	1332	GGCTGCTGCAAGGGGCTACTTTGCGTGGGCGCTTGTGCTGATTACTTGAATGCTAGATG	1391
Db	1461	GGCCCACTGCGGGGATCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1520
Qy	1392	GCTACGGAACCCAGATTCGCGGTGACGTTCAACAGACTACACACCTCAAGCGACGCCA	1451
Db	1521	GGTACACGTCGCGGTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1580
Qy	1452	AGAGTCTGCGCTGCTTCAAGAGACATGTTGGCGC	1489
Db	1581	AGAGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1618

RESULT 10
US-10-425-114-13710
; Sequence 13710, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaka, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 13710
; LENGTH: 1759
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB143-030-G7_FLI
US-10-425-114-13710

Query Match			
Best Local Similarity 51.0%; Pred. No. 1,2e-44;			
Matches 663; Conservative 1; Mismatches 600; Indels 36; Gaps 8;			
Qy	199	CACCAAGCGCGCAAGGAGTGGCTTGCATCTACACCGCTTACGATGAGACTT	258
Db	368	CGCAACCAAGCGCACCGGAGCTGACTTTGACGATACATGCTTACAGAGAGACGT	427
Qy	259	TGATCTCTTACCAATACGCGCAAGGCTTACCGCTTCTCTGCTGCTGCGGAT	318
Db	428	GATATCATGAAAGATATATGGGTTGACGCTTACCGGTTTGTGATCTCTTGTCAAGAT	487
Qy	319	CATTTCCCTCGGCGGAGCTGATCCGTCACGAGAGGAGATTAAGTTTACGCA	378
Db	488	ATTCCTCAATATGAAACCGAGAA-----GTGAATGGAAGAGTGGCTTACTACAAAG	541
Qy	379	ACTGATTGACGCGCTGTAGGCGGGGTATACGCTTGGGATGACTTGTACACTGGGA	438
Db	542	GCTGATTAATCATGTGTAAGAAAGCATCACCTTACGCAACCTGACACTACGA	601
Qy	439	TCTGCTCAGCGCTTCAAGATGCTATGAGGCTGGCTCAAGTGAAGAGGTCCAGCT	498

602 CTAACCGAGGCGCTGAGAGTCCGGTACGAGGCGCTGTTGACGACGAGAAGTGT---GAG 658
QY 499 GGAATTGAGCGGTATGAGAGTTGTGCTTGAAGTTTGGGGACCGAGTCCGAAGCTG 558
Db 659 ATCGTTGCGAGACTACCCGACCTTCTGCTTGGGGCGTTCGGCGACAGGGGGAAGACTG 718
QY 559 GATACCATCAAGACGCTGATTTAGGGCCATGATGATATGCGACCGGACAGACCGC 618
Db 719 GCTACGTTCAACAGCCGCGGGTGTGGCCGCTTGAAGGATCAACAGCGGAGGTTCCG 778
QY 619 CCCGGGAGAGGACGACATTAACAAGCACTCCACGAGGCGAACACTGCCACTGAGCCGTG 678
Db 779 GCGGGGAGGAGTGAACGGGGTCCAGG-----CCGGGGGAGGACTCGGGACCGAGCCCTA 832
QY 679 GCTGCTGGAAGAGCCCATGATGATGAGCCATGCCCCGCGCTGAGCCGTTACAGAGGA 738
Db 833 GGTGCGGCGGACCACTCATCTCTCCACGCGCGCGCTCCAGAGTACCCGCGAG 892
QY 739 CTTTGGCCCTCGCAAAAGGCGACGATCGGCATCTGCTCAACGCGGACTACTATGAGCC 798
Db 893 GCACCAAGCCAGACGAGAGGCGAGGTCGGATCTGCTGATTTCTGTGTGTAAGAGCC 952
QY 799 CTGGACAGCAATGAGCTCGGACCAAGAGGCTGCTGAGCGACGAGATGAAATTCACAT 858
Db 953 CCTCACGGCGGACTCAGCGCGCGACCGGGCGCGCTCAAGGTCCAGAGACTTCCAGT 1012
QY 859 TGGCTGTTTCCCAATCCCATCTTCTTGAAGAAGACTATCCAGAGACTGAAGAAACA 918
Db 1013 CGGATGTTCTTGAACCCCATCTGCT---ACGGGAGTACCCCAAGTCCGATCCGAGAG 1069
QY 919 GCTGGGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGCAATCTCATGTCGCGAGA 978
Db 1070 GCTCAAGGAGCGGCTCCCAAGTTCAAGGCTGAGGA---GGCCGCTTATGTCGAGGCTC 1126
QY 979 GACCGACTTCAAGGAGATTAATTAACAATCCAGTTCGCGCGCACTTGAAGCGTCC 1038
Db 1127 CATCGACTACGTCGAGAGTCAACAGTACACTGCTTACGCTGATTCAGCGGCCAA 1186
QY 1039 GGTCCCGAGAGC-----GACTATCTGGGCGCATCCATGAGACACAGAGAAATA 1089
Db 1187 CGTACGCGGCGCGCGCCAGCTACTGTCGACGTGCGACGCTGATGTTCTTGAAGC 1246
QY 1090 GGAAGCGAGCCCGCTGAGAGAGAGAGCGGCTCGCTGCGTGGCTCTGCGCGGACAT 1149
Db 1247 GCAAGGTGTGCGATTTGACCAAGGCGCACTCAGACTGCTTACATCTGCTGGGG 1306
QY 1150 GTTCCGGAAGCATCTGCGCGGAGTACGAGCTTGAAGGCAAGCCCATCTACATCAG 1209
Db 1307 ACTGTACAAAGCCGCTCACTAGTCAAGAGAAAGTACGCGCAACCCACGATGCTCTG 1366
QY 1210 GAACGAGATGCGGCTGAGAGAGAAACATGACGTGCGAGAGAGGCGCTCAACGACC 1269
Db 1367 AGAGAACGATATGAC---GACCCGCGGCAACGTCAAGTGGGCGCGGCGTGGAGAGC 1423
QY 1270 CTTCCGACATCGGTTACTTGACTCGCACTTGAATCGATTTTCAAGGCGATTACCGAGA 1329
Db 1424 GCGGAGGTTGCGCTACT---ACCGAGACTAGTCCCGACGTAAGAGGCGGAGTGAAG 1480
QY 1330 GCGGCTGCTCGTCAAGGAGTACTTGGTGGGCGTTCGATTAATCTTGAATGTTGACA 1389
Db 1481 GCGGCGCAACTGCGTGGGATCTTGGCTGCTGCTGCTGCACAACTTCGATGAGAACT 1540
QY 1390 TGGCTACGAGCCGAGATTCGGGTCACGTTCAACAGTCAACACCCCTCAGCGGAGCC 1449
Db 1541 GGGGATACGTTCCGGTTCGGCTGCTGCTACGTCAGATTCAGAGAGGCTCCGCGCTACCC 1600
QY 1450 CAAGAAGTCTGCGCTGCTCTCAAGACATGTTTCCGCGC 1489
Db 1601 CAAGAGCTCGGCTGCTGCTTCAAGGATGTATCCCGCGC 1640

RESULT 11
US-10-425-114-24850

Sequence 24850, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jindong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 24850
LENGTH: 1735
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3689-207-B2_FLI
US-10-425-114-24850

Query Match 9.3%; Score 184.6; DB 12; Length 1735;
Best Local Similarity 49.8%; Pred. No. 9.2e-44;
Matches 733; Conservative 0; Mismatches 704; Indels 36; Gaps 9;

QY 77 GCTTGGCCCAACGACTTGAATGAGGCTTTCGCAACGCGCCCTTACGAGATGAGGCGC 136
Db 166 GCGTTCCCGAGCGCTTCGTTCTTGGAGCGGCGCGTGGCGTACAGTGGAGGATG 225
QY 137 GTCAAAGAGGTGGCGCGGCCCGCTCCATCTGGGACAGTACTGCACTGAGGCCATCG 196
Db 226 GCGAAGACGCGCGGCGGGGCCCGAGCATCTGGGACCCCTTCAAGAGTTCTGGGACC 285
QY 197 GCGACCAAGCGCGCAACGCGGATGTGGCTTGGCATCTACCAACCCCTACATGAGAG 256
Db 286 ATCCCTAACATGAGCCACCGCTGACGTCAGTCAGTACAGTATCCGTACAGAGAGAT 345
QY 257 TTTGATCTTTGACCAAGTACGAGGCGCAAGGCGCTTCTGCTGTGGTCCGG 316
Db 346 GTGAACATATGAAGAACATGGCTTGTATGTCGTAACGAGTTTGTGATCTTGTGCAAG 405
QY 317 ATCATTCCTCGCGGCGAGGCTGATCCGTCACAGAGAGGAGATTTTACAGC 376
Db 406 ATTTTCCAGATGGAACCTGGAG-----GTGAACAGAGAGAGTGAATTACTACAC 459
QY 377 AAACATGTAAGCGCTGTTGAGCGGGGTATACGCTTGGGTGACTTTGTACACTGG 436
Db 460 AGGCTCATAGATTACATGCTTCACAGCAAGTATCCGCGGTATGCAAAATCTTACCATAT 519
QY 437 GATCTGCTCAGGCGCTTACGATCGCATGAGAGGCTGGCTCAACGTGAAAGGTCCAG 496
Db 520 GACTCCCATTTGGGACCTCCATGAACAGTACCTGGGCTGGCTTAC---CCAAAGATTGTG 576
QY 497 CTGACTTTGAGCGGTATGCGAGGTTGTGCTTGAAGTTTGGGAGCCGAGTCCAGAAC 556
Db 577 GAGGCGTTGAGACTAGCGCGAGTCTGCTTCCACGCGTTCCGAGACAGGTTGAAGAAC 636
QY 557 TGATCAACATCAACGAGCCCTGATTCAGGCCATCTATGATATGCCACCGGAGAAC 616
Db 637 TGGTTTACTTTCAACGAGCGGAGTGGCTGCTCTGCGCTACGAAATGGCTTGGAC 696
QY 617 GCGCGGCGGAGACAGATTAACAAGACTCCAGAGGCGACACACTGAGCGG 676
Db 697 GCACCGGAGAGT-----TTCGGGTGCGCCCGGAGGAACTCACACGAGGCG 750
QY 677 TGGCTCGCTGGAAGGCGCAGATCATGAGCCATCGCGCGCTGCGCTTACAGCAG 736
Db 751 TACTTGTGCAACATCTCATCTTCTCATGAGTGGGTGCAAGCGATACCGGAC 810
QY 737 GACTTGGCCCTCGCAAAAGGCGCAATCGGATCTTCTCAACGCGGACTACTATGAG 796

Db 811 AAGTATCAGCTTCAACGAGAGGAGATTGGAATTCCTCTGATTTCTGTGTACGAA 870
Qy 797 CCCTGGAGACGATGAGCCCTCGGAGACAGAGGCTCTGAGCAGGATGAAATTTAC 856
Db 871 CTTT--TCAGCGACAGCAATGCGACAGGCTCAGACAGCAGGACGACTTCCAC 927
Qy 857 ATTGGCTGTTTGGCAATCCCATCTTCTTGAGAGAGACTATCCAGAGCATGAGAG 916
Db 928 CTAGGCTGTTTCTTGACCCCA--TTGTACATGAGAGGTACCCGTACTGATCCAGAG 984
Qy 917 CAGTGGCGAGAGGCTTCCAGCCCTCACTCCCGGACCTTGGCATCTCAATGCCGA 976
Db 985 ATTGCCAAGACAGGCTACCGCTTGTCAAGCATGAGAGCAGAGATGTGTAAAGCTCT 1044
Qy 977 GAGACGACCTTCTACGCGCATGAAATTAACAATCCAGTTCCGCGCCACCTAGAGGT 1036
Db 1045 ATGAGCTATGTTGGCATCAACACTACACTTCTTTTACATGAGAGACCTGGGACATGG 1104
Qy 1037 CCCCTCCCGAGAGGAGATATCTCGGCGCATCCATGAGACAGAGAAATTA--GGAC 1093
Db 1105 AACCTGACGCAAGTACAGTACAGAGATGATGGCATGTGTTTGTCTACAGAGAAAT 1164
Qy 1094 GGCAGCCCGCTTGGCGAGAGAGCGGCTCGCTGGCTGGCGCTCTCGCCGAGATGTC 1153
Db 1165 GGAATTCCTATTGGCGCTCAAGCAAACTCTACTGCGCTGATGCTGTGCGTGGGAGTC 1224
Qy 1154 CGAGAGATCTCGCCCGGCTAGAGGCTCTGACGCGCAAGCCCATCTACATTA--CGAG 1210
Db 1225 AACAGGCTGTGACATATGTCAGAGAACTTACAAAAATCTCAATGATCTTGTGAA 1284
Qy 1211 AACGATGCGCCGCTCCCTGAGAGAGAAACATGACGTGCGAGAGCGCTCAAGACCC 1270
Db 1285 AACGATGAGCAACCTGTGATGTCAATTA--CAGGCTGTGATGACACA 1338
Qy 1271 TTCGCAATCCGATCTTGAATCTGCACTTGATTCGATTTCAAGGCCATTAACCCAGAC 1330
Db 1339 GTAGAAATCCGTTATTAAGAGACTAATTAAGCTCAAGAAAGCAATTA--GATGAT 1395
Qy 1331 GGCTCGCTGCAAGGGGTAATTTGGCGGCTTGTCTCGTAATCTTGGAATGTGAT 1390
Db 1396 GGTGCAAGATCATGTGGTATCTTGGCGTGTGCTGTCACACTTGAATGAGGCTT 1455
Qy 1391 GGCATGAGACCCAGATTCGCGGTCACTGACATCAACACCTCAAGCGACGCC 1450
Db 1456 GGGTACACTTGGCGTTTGGCTTGTGATGATGATCAAGATCTGAGAGAGTACCC 1515
Qy 1451 AAGAGTCTGCGCTGCTCTCAAGAGACATGTTTTCGCGCCGCGAGAGGCTTAAAGTGGC 1510
Db 1516 AAGGACTCAGCTTCTGCTTCAAGCATATGCTGTCAAGAAAGAGCTAGAGATTGCA 1575
Qy 1511 GCATTAAGAAAGGAAATTTCTTCTGATCA 1543
Db 1576 GACAAAGAGACCACTGCTCAGGTGCTATACA 1608

RESULT 12
US-10-425-114-27741
; Sequence 27741, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 27741

LENGTH: 1757
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4739-015-B1_PL1
US-10-425-114-27741
Query Match 9.3%; Score 184.6; DB 12; Length 1757;
Best Local Similarity 49.8%; Pred. No. 9.2e-44;
Matches 733; Conservative 0; Mismatches 704; Indels 36; Gaps 9;
Qy 77 GCTTCCCAAGACATTTGAATGGGCTTGGCAAGCGCCCTTACCAAGATGAAAGCGCC 136
Db 188 GCGTCCCGAGCGCTTCTGTTGAGGACGCGCGCTGACGAGTGCAGGCGATG 247
Qy 137 GTCAAGAAAGTGGCGCGCGCGCTTCAATCTGGAGACAGTACGCACTGAGGCCATTC 196
Db 248 GCAAGCAGCGCGCGCGCGCGCTTCAATGAGAGCGCTTCAATGAGGTTCTGGAGCC 307
Qy 197 GGCACCAAGCGCGCAAGCGCGATGTTGGCTTGGATCACTACCAAGCGATGAGAG 256
Db 308 ATCCCTTAACAATGCAACCGCTGACGATGACGCTGACAGAGTATCACTGTAAGAGAT 367
Qy 257 TTGATCTTGAACCAATGACGCGCAAGCGCTTACCGCTTCTTGTGTGTGCGCG 316
Db 368 GTGAACATTAATGAAGAAATGAGGCTTGTGATGCGTACCGGTTTTCGATCTCTGGT 427
Qy 317 ATCATTTCCCTCGCGCGAGCTGATCCCGTCAAGAGAGAGAAATTAAGTTTACAGC 376
Db 428 ATTTTCCAGATGGAATGGAAG--GTGAACAGAGAGAGATGATTAATCAAC 481
Qy 377 AAATGATGACGCGCTTGTGAGCGGGGTATCAAGCGCTTGGGATTTGTAACATGG 436
Db 482 AGGCTTAATTAATGATGCTTCAAGAGATGATGCGCGCTTATCAATCTTCAATAT 541
Qy 437 GATTCGCTGAGGCTTCAAGATGCTATGAGGCTGCTCAAGCTGGAAGGTCAG 496
Db 542 GACTCCCATTTGGCATTCATGAACAGTACTGGGCTGAGCTTAC--CCAAAGATTGTG 598
Qy 497 CTGACTTTAGCGGATGAGAGTGTGCTTTGAAGCTTTTGGGAGCGAGTCCAGAAC 556
Db 599 GAGCGCTTCAAGATGAGCGGATGCTTCTTCCAGCGCTTGGAGCAGGCTGAAGAAC 658
Qy 557 TGATACATCAACGAGMCCCTGATTCAGGCACTATGATATGCAACCGGAGCAAC 616
Db 659 TGGTTTACCTTCAAGCGCGAGGCTGCTGCTTGGCTTCAAGCAATGCTTGCAC 718
Qy 617 GCCCGGAGAGAGCAATTAACAGCACTCAACGAGGCAACACTGCACTGAGCG 676
Db 719 GACCGGAGAGG--TTCCGGGTGCCCCGCGAGGCAACCTCAACAGAGCGG 772
Qy 677 TGGCTGCTGAAAGCGCCAGATCATGAGCCATGCGCGCGCTGAGCTTACAGAG 736
Db 773 TACCTTGTGACACCAATCTATCTTCTTCAATGACAGCTGCGGATGATACCGGAC 832
Qy 737 GACTTGCCTCGCAAGAGCGAGATCGGATCTGCTCAACGCGCATTAATGAG 796
Db 833 AAGTATAGCTTACCGAGAGAGAGATGGAATTTCTTGGATTTCTGTGTGATGAG 892
Qy 797 CCCTGGAGACATGAGCCTCGGAGACAGAGGCTGAGCAGCAGATGAAATTTAC 856
Db 893 CTTT--TCAGCGACAGCAATGCGAGCAGGCTGACAGCAGGACGAGCACTTCCAC 949
Qy 857 ATTGGCTGTTTGGCAATCCCATCTTCTTGAGAGAGACTATCCAGAGCATGAGAG 916
Db 950 CTAGGCTGTTCTTGAACCCCA--TTGTACATGAGAGGTAACCGTCTCGATGCAAGAG 1006
Qy 917 CAGTGGCGAGAGGCTTCCAGCGCTCACTCCGCGGACTTTGCGATCTCAATGCCGA 976
Db 1007 ATTGCCAAGACAGGCTTACCGTGTGATGAGATGAGAGAGCAGATGTGAAAGCTCT 1066
Qy 977 GAGACGACTTCAAGGATGAATTAACAATCCAGTTCCGCGCCCACTAGAGGT 1036

Db 1067 ATAGACTATGTTGGATCAACCACTACTTCTTCTACATGAGAGACCCTGGACATGG 1126
Qy 1037 CCCGTCGCCGAGACGCACTATCTCGGCGCATCCATGAGACACGAGAATA---GGAC 1093
Db 1127 AACCTGAGCCGACAGTACGATACCAAGATGATTTGGCATTTGTTGTCTACAGAAAT 1186
Qy 1094 GGCACCCCGTTGGCGAGAGAGCGGCTCGCTGGCTGCGCTCTGCGCGACATGTTTC 1153
Db 1187 GGAGTTCCATTTGGCGCTCACGCAAACTCTACTGGCTGTATCATGTGGCGGAGCATC 1246
Qy 1154 CGGAGCATCTGCGCCCGGTTGATCGGCTGTACGGCAAGCCCATCTAATCA---CCGAG 1210
Db 1247 AACAAAGGCTGTAACATATGTCAAGAACTTAACAAAATCTTAACATGATCTTGTCTAA 1306
Qy 1211 AACGATGCGCCGTGCGCTGAGAGAGAAACATGACGTGAGAGAGCGCTCAACGACCC 1270
Db 1307 AACGGAATGAGACCAACCTGTGATGTCAATTAATCT-----CAGGTGTGATACACA 1360
Qy 1271 TTCCGATTCGCTACTTTGACTCGCACTTGAACCTGATTTCCAGAGCCATTAACCGAGAC 1330
Db 1361 GTAAAGATCCGTTATTAACAGAGACTAATACGACTCAAGAAAGCAATA---GATGAT 1417
Qy 1331 GCGCTGCTGCAAGGGGTAATCTTGGGCGTGTGCTGATTAATTTGAATGTGCAAT 1390
Db 1418 GGTGCAAGATCTATTTGGTACTTGTGCTGCTGCTTGAACACTTGAAGTGAAGGCTT 1477
Qy 1391 GGCCTACGAGCCGATTCGCGCTCACTTCAAGACTACACCACTTCAAGCGACGCGCC 1450
Db 1478 GGTGCTACCTGCGGCTTTGGCTGTGATACGTGATCAAGACTCTAAGAGATGATCC 1537
Qy 1451 AAGAGTCTGCGCTGTCTCAAGACATGTTTGGCGCGCGAGAGGCTTAAGTGGCG 1510
Db 1538 AAGGCTACGCTTCTGTGTTCAAGCATGCTGTCAAGAAAGAGCTAAGATGCA 1597
Qy 1511 GCATTAAGAAAGGAAATTTCTTCTTGATCA 1543
Db 1598 GACAAGAGACCACTGCGTTCAAGGTGATACA 1630

RESULT 13
US-10-369-493-31938
; Sequence 31938, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31938
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Thermobifida fusca
US-10-369-493-31938

Query Match 9.2%; Score 183.2; DB 15; Length 1338;
Best Local Similarity 51.7%; Pred. No. 2.1e-43;
Matches 714; Conservative 1; Mismatches 589; Indels 78; Gaps 10;

Qy 92 TTTAAATGGGCTTCGCAACGCGCGCTACCAATGCAAGCGCGCTCAAGAAAGTGGC 151
Db 1 TTCTGTGGGAGTGGGACCGCTTCGTTCCAGATCAAGGCTTCACCAACGCGCGACGCG 60
Qy 152 CGCGCCGCTCATCTGGGACACGTAACCTGAGGACATGCGGCAACCAACGCGCGC 211
||||| ||||||| ||||| | |||||

Db 61 CGCGCCCAAGACATCTGGGACACTTCTGGGCCACTCCCGGCAAGTGCAGAAACGCGAC 120
Qy 212 AACGCGATGTGGCTTCCGATCACTACACCGCTACGATAGAGACTTGTATCTTACC 271
Db 121 ACGGCGACCCCTGCTCGACACACTACACCGGATACCGCATATACGAGGCTGTATCCG 180
Qy 272 AAGTACCGCCAAAGGCTTACCGCTTCTCTTGTGCGGTGCGGATCATTTCCCTCGG 331
Db 181 GAGTGGGCGTGGGCGCTACCGCTTCTCATGCGCTGCGCGGATACAGCCCGAGGGC 240
Qy 332 GCGAGCTGTATCCGTTCAAGAGAGGAATTGATTTTACAGCAAACTGATTAAGCGC 391
Db 241 -----AAGGCAACGCGCTGGAGGCGGCGCTGACTTACACCGGCTTGTGAGTGC 294
Qy 392 CTGTGAGGGGGGTATCACGCTTGGGTGATCTTTGATACATGAGGATCTGAGGCG 451
Db 295 CTGCTGAGGCGGCACTGAGCGGTGCGCAACCTTACCATGAGGACTGCGCGAGGCG 354
Qy 452 CTTCACGATGCTATGAGGCTGCTCAACGTGGAAGGTCCAGCTGACTTTGAGCGG 511
Db 355 CTGAGAGA---CGCGCGGCTGCGCAACCGGGAACCGGCCAAGCG---TTCCGCGAC 408
Qy 512 TATGCAAGTGTGCTTTGAACGTTTGGGACCGAGTCCAGACTGATACATCAAC 571
Db 409 TACGCGAGATGCTTACCGCGGCTCGGCGACCGGATCCAACTGGAACACGCTAAC 468
Qy 572 GAWCCCTGATTAAGGCTCTATGATATGCAACCGGCGCAACCGCCCGGCGAGAGC 631
Db 469 GAGCGTGTGCTCGGCTTCTGAGCTTCCGCGGTGCAACGCGCGCGCGC----- 523
Qy 632 AGCATTAACAAGACTTCAACCGAGGCAACACTGAGCGCTGAGCGCTGAGAAAG 691
Db 524 -----GCAGAGACCGGCTGCTGCGTGGCGCGCGCCAC 558
Qy 692 GCCCAATCATGAGCCATGCGCGCGCTGAGCGCTTCAACAGAGGACTTTGCGCCCTG 751
Db 559 CACCTGATGTGGGCAACCGGCTGCGCGCTGCGCTG---ATGCGGAGACTTGGCGGCGCAG 615
Qy 752 CAAGAGGCGCAGATGCGCATCTGCTCAACGCGGCACTATGATAGGCTTGGGACAGCAAT 811
Db 616 GCGGAGCTTCCGTGGGATGCTGTGCGGCAACCAACGACCGGTCCGTCTTACACT 675
Qy 812 GAGCTTGGGACAAAGAGCTGTGAGCGACGAGTGAATTTCAACTTGGCTTGTGCG 871
Db 676 GACAGTGAAGCGACCGGAGCGCTGCGCGGATTTGACGCGCTTGGGAACCGCATCTTC 735
Qy 872 AATCCATCTTCTTGAAGAAAGACTATCCAGAGAGCATGAAGAGAGCTGGCG--- 926
Db 736 ACCGAGCGCTGTGAAGGCGCTTACCGGAGAGCTTATCGAGGACGTCGCGCGGCTC 795
Qy 927 -----AGAGCTTCCAGCCCTCACTCCGCGGA-----CTTTCGCAATCTCAATGCC 973
Db 796 ACCGACTACAGCTTCTCTCAGAGCGGCACTGAAGACATCTCCGCAACTGAGCAG 855
Qy 974 GGAAGAACGACTTCTACCGCATGAATTAATAACATCCAGTTCGCGCGCACTTAAGC 1033
Db 856 ATGGGCGTCAACTTCAACACCGAGCTGGTGTGAGCAACCGGGAAGAACGCGGCTCC 915
Qy 1034 GATCCGCTCCGAGAGAGACTATCTCGGCGCATCAATGAGACCAAGAGAAATAGAGC 1093
Db 916 GACCGGCTGCGGACGAGGACTTACTCGCGGTGAT---GGAGAGAGATGTCTGTGAG 972
Qy 1094 GCGAGCCCGTTGGCGAGAGAGCGGCTTGGCTGCTGCTGCGCGGACATGTTTC 1153
Db 973 GTGAGACCCGCGCTGCGGTGAGACCGGCAAGGGCTGAGTGAACCCGAGCCGCGTATC 1032
Qy 1154 CGAAGCATCTGCGCGGCTGTACGCGCTGTACGCGCAACCCATCTACACCGAAGC 1213
Db 1033 GACACGCTGACCCGCGTGGCAACGACTACCCGCGCTCCGCTGTACATCACCGAAGC 1092
Qy 1214 GATGCGCGGCTTGAAGAGAAACATGAGTGTGAGAGAGCGCTCAACCCCTTC 1273
Db 1093 GCGCGCGCTTC-----GAGGCAAGGTGTGAGCGCGGCTGACGACACCGAG 1143


```

; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kieps, Joel
; APPLICANT: Zhu, Tong
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1282
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1282

```

Query Match 9.2%; Score 182.8; DB 11; Length 1734;

Best Local Similarity 49.2%; Pred. No. 3.1e-43;

Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

```

OY 79 TCTGCCCAACGACTTTGAATGGGCTTGCAACGGCCCTACAGATCCGAAGCGCCGT 138
DB 90 TTTTCTCGATGATTTTCATCTTTGGACAGCTGCTCCCTGGCGTTTCAAGTACGAGTGAAC 149
OY 139 CAAGAAGGTGGCGCGCGCCGCTCCATCTGGGACACGTAACCTGACCTGAGCCATCGCG 198
DB 150 AAGTGAAGGTGGCAAGTCTTCACTATATGGGATCCTTCAAGCTTCAAGTATCCAGAAAG 209
OY 199 CACCAACGGCGCCCAACGGCGATGTGGCTTGGATCACTACCAACCGCTACGATGAGACTT 258
DB 210 GACCAAAATGATATGAGATGAGATGAGCAATGATTTTATCATGCTTCAAGGATGACAT 269
OY 259 TGATCTCTTGAACAGTACGGCGCAAGCCCTACCGCTTCCCTTGTGCTGCTGCGGAT 318
DB 270 AAAATTGATGAGAGGCTAAACATGACGCTTCCGATTTTCAACTCTGCTGATCAAGATT 329
OY 319 CATTCGCCCTCGCGCGCAGGCTGATCCCGTCAACGAGGAGGAAATTGATTTTACAGCA 378
DB 330 AATACCGAGTGAAGATGAAGATGAGTAAACAAAGATGATCAATTTCAAGAGGA 389
OY 379 ACTGATGACCGCTGTTGAGGCGGGGTATCAAGCTTGGGATCTTGTACCACTGGGA 438
DB 390 TCTCATGACGAACTTTGCTTGTGCTAATGACATCAACCTTCGATGACGCTCTATCATTTGGGA 449
OY 439 TCTGCTCAGGCGCTTCAGATCGCTATGAGGCGTGGCTCAACGTAAGAGGTCACAGT 498
DB 450 CCACCCCAATCTTTGAGAGAGCAAAATGATGCTTTCTAAAGCCTTAATTCGTAAGA--- 506
OY 499 GGAATTTGAGCGGTATGAGAGGTTTGTCTTGAACGTTTGGGGAACGATCCAGACTG 558
DB 507 AGATTTTGAAGATTTTGAAGATTTGTTTGAAGATTTGGAAGATTAAGATGATG 566
OY 559 GATCAACATCAACGAAACCTGATTCAGGCAATCTATGATATGCAACCGGAGCAAGC 618
DB 567 GACAACATCAACGAAACCTTATATATGATGCTTGGGTTATGATCAAGGTAAACAGGC 626
OY 619 CCGCGGAGGAGC-----AGCATTAACAAGCACTCAACCGAGGCAACACTGCACTGA 672
DB 627 GAGTGAAGATGCTCAAAATGAGTAAAGAAAAGTGTAGGCTGAAGATTCAGATCCGA 686
OY 673 GCGGTGCTCGCTGGAAGGCCAGATCATGAGCCATGCGCGCTGCGCTCTACAG 732
DB 687 GCTTACATGTTTTCACATCACTCTTGTGCTGCTGCTGCAAGTGAAGATTTG 746
OY 733 CAGGGAATTTTGGCGCGGCAAAAGGCGCAATCGGCAATCTGCTCAACGCGGACTACTGA 792

```

```

DB 747 AAAATGTAAGAAAACTTCGATGATGCGCAAAATTTGGATATGATCATCAACCAAGATGTT 806
OY 793 TGAGCCCTGGGACAGCAATGAGCTTCGGGACAGAGAGCTGTAGCGACGATGGAATT 852
DB 807 CGAGCTTATCATTTCCGATTTCAACTGATTAAGAGAGCTGGAAGAGCTTGTGCTT 866
OY 853 TCACATTTGCTGTTTGGCAATCCATCTTCTTGAAGAGACTATTCAGAGCATGAA 912
DB 867 TGAATTTGATGCAATCTTGAATCAGATTC---ACGAGATTAATCAGAGATTTGAAA 923
OY 913 GAAGCACTGGGAGAGAGGTTCCAGCCCTCACTCCCGCGGACTTTGGCATTCCTCAATGC 972
DB 924 AAGTACGCGGAAATTAATTAATCTTCTTCTTGAACCAATCAAGATTTACAAA 983
OY 973 CGGAGAGACCGACTTCTACGCGCATGATTAATCAATCCAGTTGCGCGCCACTAGA 1032
DB 984 TTCA---TCAGATTTGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1040
OY 1033 CGGTCCCGTCCCGAGACGAGCTATCTGGCGCCATTCATGA-----GCAACAGGA 1083
DB 1041 TCACATGACCTCGAATAACCTGTTCAAACTGACACACATGAGATGAAATGAC 1100
OY 1084 GAATTAAGAGAGGAGCGCCGTTGGCGAGAGAGCGGCTGCGTGGCTGCGCTCGGCC 1143
DB 1101 TAAATCAAGTGTCAATCAATCGAAGCTTGGGAAAGAGGCTTCTTATTTTCAACACC 1160
OY 1144 GGAATGTTCCGAGAGCATCTGCGCGGTGTACGGCTGTAC---GGCAAGCCATCTA 1200
DB 1161 GGAAGCTTACGAAAGTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1220
OY 1201 CATCAACGAGAGAGATGCGCGTCCCTGAGAGAGAGAAATATGACGTCGAGAGAGGCTT 1260
DB 1221 CATCAAGAAATGAAATTAATGACACGACGATCAAAACCAAGAGAAATTTGT 1280
OY 1261 CAAGACCCCTTCGCAATCTGCGGTATCTGACTGCACTTGAATTCGATTTCCAAGGCAT 1320
DB 1281 GAAGAGACATTTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1340
OY 1321 TACCAGAGAGCGCTGCTGCTCAAGGGATCTTGGCTGGGCGTTGCTGATTAATCTTGA 1380
DB 1341 AGTGAAGATGAGGTGACGTAAGAGATTAATGCAATGATCATTAATTAATTAATTAAT 1400
OY 1381 ATGTCAGATGCTACGGAACCCAGATTCGG 1410
DB 1401 ATGGAGACATGATACACTGCAAGATTTGG 1430

```

Search completed: March 29, 2004, 03:16:31
Job time : 706.681 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 23:31:59 ; Search time 6860.46 Seconds
(without alignments)
1021.501 Million cell updates/sec

```

Title:      US-10-026-140-1
Perfect score: 1991
Sequence:   1 agccacgtcgcaaccagcag.....tttlaaaaaaaaaaaaaaa 1991

```

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues
Total number of hits satisfying chosen parameters: 75154660

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database

1:	/cgn2_6/prodata/2/pna/PCnts_COMB_seg.*
2:	/cgn2_6/prodata/2/pna/US06_COMB_seg.*
3:	/cgn2_6/prodata/2/pna/US07_COMB_seg.*
4:	/cgn2_6/prodata/2/pna/US080_COMB_seg.*
5:	/cgn2_6/prodata/2/pna/US081_COMB_seg.*
6:	/cgn2_6/prodata/2/pna/US082_COMB_seg.*
7:	/cgn2_6/prodata/2/pna/US083_COMB_seg.*
8:	/cgn2_6/prodata/2/pna/US084_COMB_seg.*
9:	/cgn2_6/prodata/2/pna/US085_COMB_seg.*
10:	/cgn2_6/prodata/2/pna/US086_COMB_seg.*
11:	/cgn2_6/prodata/2/pna/US087_COMB_seg.*
12:	/cgn2_6/prodata/2/pna/US088_COMB_seg.*
13:	/cgn2_6/prodata/2/pna/US089_COMB_seg.*
14:	/cgn2_6/prodata/2/pna/US090_COMB_seg.*
15:	/cgn2_6/prodata/2/pna/US091_COMB_seg.*
16:	/cgn2_6/prodata/2/pna/US092b_COMB_seg.*
17:	/cgn2_6/prodata/2/pna/US093a_COMB_seg.*
18:	/cgn2_6/prodata/2/pna/US093b_COMB_seg.*
19:	/cgn2_6/prodata/2/pna/US094_COMB_seg.*
20:	/cgn2_6/prodata/2/pna/US095a_COMB_seg.*
21:	/cgn2_6/prodata/2/pna/US095b_COMB_seg.*
22:	/cgn2_6/prodata/2/pna/US096_COMB_seg.*
23:	/cgn2_6/prodata/2/pna/US095c_COMB_seg.*
24:	/cgn2_6/prodata/2/pna/US095d_COMB_seg.*
25:	/cgn2_6/prodata/2/pna/US096a_COMB_seg.*
26:	/cgn2_6/prodata/2/pna/US096b_COMB_seg.*
27:	/cgn2_6/prodata/2/pna/US096c_COMB_seg.*
28:	/cgn2_6/prodata/2/pna/US096d_COMB_seg.*
29:	/cgn2_6/prodata/2/pna/US096e_COMB_seg.*
30:	/cgn2_6/prodata/2/pna/US097a_COMB_seg.*
31:	/cgn2_6/prodata/2/pna/US097b_COMB_seg.*
32:	/cgn2_6/prodata/2/pna/US097c_COMB_seg.*
33:	/cgn2_6/prodata/2/pna/US098a_COMB_seg.*
34:	/cgn2_6/prodata/2/pna/US098b_COMB_seg.*
35:	/cgn2_6/prodata/2/pna/US098c_COMB_seg.*
36:	/cgn2_6/prodata/2/pna/US098d_COMB_seg.*
37:	/cgn2_6/prodata/2/pna/US099a_COMB_seg.*
38:	/cgn2_6/prodata/2/pna/US099b_COMB_seg.*
39:	/cgn2_6/prodata/2/pna/US099c_COMB_seg.*
40:	/cgn2_6/prodata/2/pna/US099d_COMB_seg.*
41:	/cgn2_6/prodata/2/pna/US099e_COMB_seg.*
42:	/cgn2_6/prodata/2/pna/US100f_COMB_seg.*
43:	/cgn2_6/prodata/2/pna/US100a_COMB_seg.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

QY 1021 GCGCCACTAGACGGTCCCGTCCCGAGAGACGACTATCTGGCGCATTCATGACACCA 1080
DB 1021 GCGCCACTAGACGGTCCCGTCCCGAGAGACGACTATCTGGCGCATTCATGACACCA 1080
QY 1081 GAGAAATAGAGACGGCGCCCGTGGCGAGAGAGCGCGCTCGCTGGCTGGCTCTG 1140
DB 1081 GAGAAATAGAGACGGCGCCCGTGGCGAGAGAGCGCGCTCGCTGGCTGGCTCTG 1140
QY 1141 CCGGACATCTCCGGAAGATCTCGCGCGGTGTAGCGCTGTAGCGGCAAGCCATCTA 1200
DB 1141 CCGGACATCTCCGGAAGATCTCGCGCGGTGTAGCGCTGTAGCGGCAAGCCATCTA 1200
QY 1201 CATACCGAGAGACGATGCGCGTGGCGAGAGAGACGATGACGTCGAGAGAGCGCT 1260
DB 1201 CATACCGAGAGACGATGCGCGTGGCGAGAGAGACGATGACGTCGAGAGAGCGCT 1260
QY 1261 CAAGACACCCCTCCGATCCGCTGCTTGAATGCACTGGCACTGATTCGAAAGCCAT 1320
DB 1261 CAAGACACCCCTCCGATCCGCTGCTTGAATGCACTGGCACTGATTCGAAAGCCAT 1320
QY 1321 TACCCAGAGAGCGGTGCTGCTCAAGGGGTACTTGGCGTGGCGCTGCTGATTAATGGA 1380
DB 1321 TACCCAGAGAGCGGTGCTGCTCAAGGGGTACTTGGCGTGGCGCTGCTGATTAATGGA 1380
QY 1381 ATGGTCAGATGCTACGAGACCGATTCGCGCTCAAGATCAAGATCAAGATCAAGATCA 1440
DB 1381 ATGGTCAGATGCTACGAGACCGATTCGCGCTCAAGATCAAGATCAAGATCAAGATCA 1440
QY 1441 GCGCAGCGCCAGAGATGCTGCTGCTCAAGAGATGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 GCGCAGCGCCAGAGATGCTGCTGCTCAAGAGATGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 TAAAGTGGCGGCTTAAAGGAAATTTCTTGGATTCAGACCTGATGATCTTCC 1560
DB 1501 TAAAGTGGCGGCTTAAAGGAAATTTCTTGGATTCAGACCTGATGATCTTCC 1560
QY 1561 TCTCTCTTCTCCCTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
DB 1561 TCTCTCTTCTCCCTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY 1621 CCCCCGCTTCTCTCATGACCCCATGCTCTTGGCCCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1621 CCCCCGCTTCTCTCATGACCCCATGCTCTTGGCCCTGCTGCTGCTGCTGCTGCTGCT 1680
QY 1681 GCTCTTATACGCTTATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
DB 1681 GCTCTTATACGCTTATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1741 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
DB 1741 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
QY 1801 AGCACAAGTTCAAGACGAGACGAGACGAGACGAGACGAGACGAGACGAGACGAGACG 1860
DB 1801 AGCACAAGTTCAAGACGAGACGAGACGAGACGAGACGAGACGAGACGAGACGAGACG 1860
QY 1861 AAGAGCCATGTTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
DB 1861 AAGAGCCATGTTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
QY 1921 CCGAGGAGAGAGGCAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
DB 1921 CCGAGGAGAGAGGCAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
QY 1981 AAAAAAAAAA 1991
DB 1981 AAAAAAAAAA 1991

RESULT 2
PCT-US02-34764-1
; Sequence 1, Application PC/TUS0234764

GENERAL INFORMATION:
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Goedegebuur, Frits
APPLICANT: Ward, Michael
APPLICANT: Yao, Jian
TITLE OF INVENTION: BGLs Beta-Glucosidase and Nucleic Acids
FILE REFERENCE: GC697
CURRENT APPLICATION NUMBER: PCT/US02/34764
CURRENT FILING DATE: 2003-10-30
NUMBER OF SEQ. ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ. ID NO 1
LENGTH: 1991
TYPE: DNA
ORGANISM: Trichoderma reesei
PCT-US02-34764-1

Query Match 100.0%; Score 1990.2; DB 1; Length 1991;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1991; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGGTCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60
DB 1 AGCAGGTCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60
QY 61 AATGCCGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
DB 61 AATGCCGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 121 CCGATGCAAGGCGCGCTCAAGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 180
DB 121 CCGATGCAAGGCGCGCTCAAGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 180
QY 181 CCACTGAGACCATTCGCGCAACGCGCGCAACGCGCGATGCGCTGCGATCACTACCA 240
DB 181 CCACTGAGACCATTCGCGCAACGCGCGCAACGCGCGATGCGCTGCGATCACTACCA 240
QY 241 CCGCTAGATGAGACCTTGTATCTTGAACGATGCGCGCAAGGCGCTGCGATCACTACCA 300
DB 241 CCGCTAGATGAGACCTTGTATCTTGAACGATGCGCGCAAGGCGCTGCGATCACTACCA 300
QY 301 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 301 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 361 AATTGAGTTTACAGCAAACTGATGACGCGCTGTTGAGCGGGGTATACGCTTGGGT 420
DB 361 AATTGAGTTTACAGCAAACTGATGACGCGCTGTTGAGCGGGGTATACGCTTGGGT 420
QY 421 GACTTGTACCACTGGGATCTGCTGAGCGCTTCAAGATGCTTATGAGAGGCTGGCTCAA 480
DB 421 GACTTGTACCACTGGGATCTGCTGAGCGCTTCAAGATGCTTATGAGAGGCTGGCTCAA 480
QY 481 CGTGGAAGAGTCACTGAGCTTGAAGCGGTATGAGAGGTTGCTTGAAGCTTTGG 540
DB 481 CGTGGAAGAGTCACTGAGCTTGAAGCGGTATGAGAGGTTGCTTGAAGCTTTGG 540
QY 541 GAGCGAGTCAGAACTGATCAATCAACGAGCCCTGATTCAGGCTATGATGATA 600
DB 541 GAGCGAGTCAGAACTGATCAATCAACGAGCCCTGATTCAGGCTATGATGATA 600
QY 601 TGGCAGCGGAGCAAGCG 660
DB 601 TGGCAGCGGAGCAAGCG 660
QY 661 CACTGCACTGAGCGCTGCTGCTGGAAGGCGCGAGATCATGAGCATGCGCGCGCT 720
DB 661 CACTGCACTGAGCGCTGCTGCTGGAAGGCGCGAGATCATGAGCATGCGCGCGCT 720
QY 721 GCGCGTCTACAGGAGGACTTTGCGCCCTGCAAAAGGCGCGAGATGCGCATCTGCTCAA 780
DB 721 GCGCGTCTACAGGAGGACTTTGCGCCCTGCAAAAGGCGCGAGATGCGCATCTGCTCAA 780

[illegible]

QY	1861	AAAGCCCATGTTAGCCTCATGTTACATGCTATGTAATCTACATCAGCCATTCACTA	1920
Db	1861	AAAGCCCATGTTAGCCTCATGTTACATGCTATGTAATCTACATCAGCCATTCACTA	1920
QY	1921	CCAGGCCAAGGGCCACAGAGAGTCTCATGCTCTTACTCTGATAATAAGCGTTTTTAAAA	1980
Db	1921	CCAGGCCAAGGGCCACAGAGAGTCTCATGCTCTTACTCTGATAATAAGCGTTTTTAAAA	1980
QY	1981	AAAAAAAAAA	1991
Db	1981	AAAAAAAAAA	1991

```

RESULT 3
US-10-026-140-1
Sequence 1, Application US/10026140
GENERAL INFORMATION:
APPLICANT: Dunn-Coleman, Nigel
INVENTOR: Dunn-Coleman, Nigel
APPLICANT: Goedegebuurt, Frits
INVENTOR: Goedegebuurt, Frits
APPLICANT: Yao, Jian
INVENTOR: Yao, Jian
TITLE OF INVENTION: BGIs Beta-Glucosidase
FILE REFERENCE: Encoding the Same
FILE REFERENCE: GCG697
CURRENT APPLICATION NUMBER: US/10/026,140
CURRENT FILING DATE: 2002-03-26
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1991
TYPE: DNA
ORGANISM: Trichoderma reesei
US-10-026-140-1

```

Query Match	100.0%	Score 1990.2	DB 43	Length 1991
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1991; Conservative	0	Mismatches	0	Gaps 0

[illegible]


```

Qy 541 GACCGAGTCCAGAACTGATACCATCAAGAACCTGTGATTCAGGCCATCTATGATA 600
Db 541 GACCGAGTCCAGAACTGATACCATCAAGAACCTGTGATTCAGGCCATCTATGATA 600
Qy 601 TGCACCGGAGCAACGCGCGGAGAGAGAGATTAACAGAGACTTCAACGAGGCA 660
Db 601 TGCACCGGAGCAACGCGCGGAGAGAGAGATTAACAGAGACTTCAACGAGGCA 660
Qy 661 CACTGCACTGAGAGCCGTGCTGCTGGAAGGCCAATCATGAGCCATGCGCGCGCT 720
Db 661 CACTGCACTGAGAGCCGTGCTGCTGGAAGGCCAATCATGAGCCATGCGCGCGCT 720
Qy 721 GGCCTGTACAGAGAGGACTTTCGCGCCCTCGCAAAAGGCGAGATTCGAGCTCA 780
Db 721 GGCCTGTACAGAGAGGACTTTCGCGCCCTCGCAAAAGGCGAGATTCGAGCTCA 780
Qy 781 CGCGCATCTACTATGAGCCCTGCGGAGAGCAATGAGCCCTCGGAGCAAGAGGCTG 840
Db 781 CGCGCATCTACTATGAGCCCTGCGGAGAGCAATGAGCCCTCGGAGCAAGAGGCTG 840
Qy 841 ACCGATGGAATTTTACATTTGGCTGTTTGCATATCCCATCTTCTTGAAGAGACTATCC 900
Db 841 ACCGATGGAATTTTACATTTGGCTGTTTGCATATCCCATCTTCTTGAAGAGACTATCC 900
Qy 901 AGAGAGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Db 901 AGAGAGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Qy 961 CATCTCTAATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 961 CATCTCTAATGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Qy 1021 GCGCCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 1021 GCGCCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy 1081 GAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 GAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy 1141 CCGGAGCATGTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1141 CCGGAGCATGTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Qy 1201 CATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 CATCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1261 CAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 CAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 1321 TACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 TACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy 1381 ATGCTCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 ATGCTCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy 1441 GCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1441 GCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Qy 1501 TAAAGTGGCGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 1501 TAAAGTGGCGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Qy 1561 TCTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Db 1561 TCTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620

```

```

Qy 1621 CCCCCGCTTCTTCTCATGACCCCATGCTCTTGCCCTTG6CCCTCTCTGCGAATCT 1680
Db 1621 CCCCCGCTTCTTCTCATGACCCCATGCTCTTGCCCTTG6CCCTCTCTGCGAATCT 1680
Qy 1681 GCTCTTATCAGGCTTATGCGTGTGTTTACTTGGCTTTTCTTTTCTCTTCTGT 1740
Db 1681 GCTCTTATCAGGCTTATGCGTGTGTTTACTTGGCTTTTCTTTTCTCTTCTGT 1740
Qy 1741 CTGTCTGTGCGCTGTATGCTATGCTATCTGAGCCCTTGCTCATTTGGAGAGATCT 1800
Db 1741 CTGTCTGTGCGCTGTATGCTATGCTATCTGAGCCCTTGCTCATTTGGAGAGATCT 1800
Qy 1801 AGCACAAGTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1860
Db 1801 AGCACAAGTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1860
Qy 1861 AAAGCCCATGTTTAAAGCTTATGTTTCAATTTGTTATTTATCATCAGCCATTCCTA 1920
Db 1861 AAAGCCCATGTTTAAAGCTTATGTTTCAATTTGTTATTTATCATCAGCCATTCCTA 1920
Qy 1921 CCAAGGCAAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 CCAAGGCAAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy 1981 AAAAAAAAAA 1991
Db 1981 AAAAAAAAAA 1991

```

```

RESULT 4
PCT-US02-34674-3
; Sequence 3, Application PC/RUS0234674
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: PCT/US02/34674
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Trichoderma reesei
PCT-US02-34674-3

```

```

Query Match 73.0%; Score 1454.2; DB 1; Length 1455;
Best Local Similarity 100.0%; Pred. No. 2,2e-300;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 62 ATGCGGAGTCTGAGTCTGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
Db 1 ATGCGGAGTCTGAGTCTGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Qy 122 CAGATCGAAGGCGCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181
Db 61 CAGATCGAAGGCGCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Qy 182 CACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
Db 121 CACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Qy 242 CGCTACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
Db 181 CGCTACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Qy 302 TTGTCTGTGTCGCGGATCATTTCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
Db 241 TTGTCTGTGTCGCGGATCATTTCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

```


QY	362	ATTGAGTTTTCACGAAACGTGATTGACGCGCTGTTGAGGCGGGATATCACGCTTTGGGTG	421
Db	301	ATTGAGTTTTCACGAAACGTGATTGAGCGCCTGTTGAGGCGGGTATCACGCTTTGGGTG	360
QY	422	ACTTTGTAACCACTGGGATCTGCGCTCAGGCGTTTACGATGTGCTATGGAAGCTGGCTCAAC	481
Db	361	ACTTTGTAACCACTGGGATCTGCGCTCAGGCGTTTACGATGTGCTATGGAAGCTGGCTCAAC	420
QY	482	GTGGAAGAGTCCAGCTGAGACTTTAGCGGATATGCGAGTTGTGCTTTGAAAGCTTTGGG	541
Db	421	GTGGAAGAGTCCAGCTGAGACTTTAGCGGATATGCGAGTTGTGCTTTGAAAGCTTTGGG	480
QY	542	GACGAGTCCAGAACTGGATACCATTCATAAGAACCTTGGATTCAAGCCATCTATGATAT	601
Db	481	GACGAGTCCAGAACTGGATACCATTCATAAGAACCTTGGATTCAAGCCATCTATGATAT	540
QY	602	GCCACCGGCAGCAACCGCCCCGGGACAGAGCAGATTAAACAAGCACTCAACCGAGGCAC	661
Db	541	GCCACCGGCAGCAACCGCCCCGGGACAGAGCAGATTAAACAAGCACTCAACCGAGGCAC	600
QY	662	ACTGCACTGAGCGCGTGGCTCGCTGGAAAAGCCCAATCATGAGCCATCCCGCGCGTG	721
Db	601	ACTGCACTGAGCGCGTGGCTCGCTGGAAAAGCCCAATCATGAGCCATCCCGCGCGTG	660
QY	722	GCCGCTTACAGCAGGAGACTTTGCGCCCTCGCAAAAAGGCGAGATCGGCATCTCGCTCAAC	781
Db	661	GCCGCTTACAGCAGGAGACTTTGCGCCCTCGCAAAAAGGCGAGATCGGCATCTCGCTCAAC	720
QY	782	GGCGATTACTATGAGCCCTTGGGACAGCAATGAGCCTCGGAGACAAGAGGCTGCTGAGCA	841
Db	721	GGCGATTACTATGAGCCCTTGGGAGCAAGCAATGAGCCTCGGAGACAAGAGGCTGCTGAGCA	780
QY	842	CGATGGAATTTTCACTTGGCTGTTTGGCAATCCCATCTTCTTAAAGAAAGCATATCA	901
Db	781	CGATGGAATTTTCACTTGGCTGTTTGGCAATCCCATCTTCTTAAAGAAAGCATATCA	840
QY	902	GAGAGCATGAAGAAGCAGCTGGGCGAGAGGCTTTCAGGCGCTTCACTCCGCGGACTTTGCC	961
Db	841	GAGAGCATGAAMAAACAGCTGGGCGAGAGGCTTTCAGGCGCTTCACTCCGCGGACTTTGCC	900
QY	962	ATCTCTCAATGCGGAGAGACCGACCTCTTACCGGCAAGAAATTACTACATCCCATGTTGCGG	1021
Db	901	ATCTCTCAATGCGGAGAGACCGACCTCTTACCGGCAAGAAATTACTACATCCCATGTTGCGG	960
QY	1022	CGGCACCTBAGACGATCCCGTCCCGAGACGCACTATCTCGGCGCCATCATAGACACAG	1081
Db	961	CGGCACCTBAGACGATCCCGTCCCGAGACGCACTATCTCGGCGCCATCATAGACACAG	1020
QY	1082	GAGATTAAGAAGCGGACGCGCGTGGCGAGAGAGCGGCGCTTGCTGCTGCTGCTGCG	1141
Db	1021	GAGATTAAGAAGCGGACGCGCGTGGCGAGAGAGCGGCGCTTGCTGCTGCTGCTGCG	1080
QY	1142	CCGACATGTTCCGGAAACATCTCGCCCGGGGTATACGGCTGTACCGGCAAGCCCATCTAC	1201
Db	1081	CCGACATGTTCCGGAAACATCTCGCCCGGGGTATACGGCTGTACCGGCAAGCCCATCTAC	1140
QY	1202	ATCACCGGAAGGATGCCCCGTGCGCTTGGAGAAGAGAAATGATGCTGCGAGAGAGCGCTC	1261
Db	1141	ATCACCGGAAGGATGCCCCGTGCGCTTGGAGAAGAGAAATGATGCTGCGAGAGAGCGCTC	1200
QY	1262	AAGCAACCCCTTCGCGATCCGCTACTTTGACTCGACACTTGGACTGTGATTTCCAAAGGCATT	1321
Db	1201	AAGCAACCCCTTCGCGATCCGCTACTTTGACTGTGACTGTGATTTCCAAAGGCATT	1260
QY	1322	ACCCAGAGACGGCGTCTGCTGAAGGGTACTTTGCGTGGCGCTTGCTGATTAACCTTGGAA	1381
Db	1261	ACCCAGAGACGGCGTCTGCTGAAGGGTACTTTGCGTGGCGCTTGCTGATTAACCTTGGAA	1320
QY	1382	TGGTCAATGAGCTTACGAGACCAAGTTTGGGGTCAAGTTTCAAGACTTACACCACTTCAAG	1441
Db	1321	TGGTCAATGAGCTTACGAGACCAAGTTTGGGGTCAAGTTTCAAGACTTACACCACTTCAAG	1380

Query	1442	CGACACGCCAAGAAAGCTCCCTCGTTCCTCAAGGACATGTTTGGCGCCCGGACAGGGTT	1501
Db	1381	CGACACGCCAAGAAAGCTCCCTCGTTCCTCAAGGACATGTTTGGCGCCCGGACAGGGTT	1440
Query	1502	AAAGTGCGGCATPAA	1516
Db	1441	AAAGTGCGGCATPAA	1455

RESULT 5

PCT-US02-34764-3

; Sequence 3, Application PC/TUS0234764

; GENERAL INFORMATION:

; APPLICANT: Dunn-Coleman, Nigel

; APPLICANT: Goedegebuurt, Frits

; APPLICANT: Ward, Michael

; APPLICANT: Yao, Jian

; TITLE OF INVENTION: BGIs Beta-Glucosidase and Nucleic Acids

; TITLE OF INVENTION: Encoding the Same

; FILE REFERENCE: GC697

; CURRENT APPLICATION NUMBER: PCT/US02/34764

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 1455

; TYPE: DNA

; ORGANISM: Trichoderma reesei

PCT-US02-34764-3

Query Match	73.0%; Score 1454.2; DB 1; Length 1455;		
Best Local Similarity	100.0%; Pred. No. 2.2e-300;		
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Query	62	ATGCCCGAGTGGCTAGCTCTGCCCAAGACTTTGAATGGGGCTTCGCAACGGCCGCTTAC	121
Db	1	ATGCCCGAGTGGCTAGCTCTGCCCAAGACTTTGAATGGGGCTTCGCAACGGCCGCTTAC	60
Query	122	CAGATCGAAGGCGCCGTCAAAAGAGTGGCGCGCCCGCTTCATCTGGGACACGTACTGC	181
Db	61	CAGATCGAAGGCGCCGTCAAAAGAGTGGCGCGCCCGCTTCATCTGGGACACGTACTGC	120
Query	182	CACCTGGAGCCATGGCGGCACCAACGGGCGCAACGGCGATGTGGCTTGCATCACTACAC	241
Db	121	CACCTGGAGCCATGGCGGCACCAACGGGCGCAACGGCGATGTGGCTTGCATCACTACAC	180
Query	242	CGCTACGATGAGACTTTGATCTCTTGAACCAAGTACGGGCGCAAAAGGCCCTACCCGCTTCC	301
Db	181	CGCTACGATGAGACTTTGATCTCTTGAACCAAGTACGGGCGCAAAAGGCCCTACCGCTTCTCC	240
Query	302	TTGTCTGTGTGTCGGGATCATTTCCCTCTGGCGGCGAGGCTGATCCGTCACGAGAGGGA	361
Db	241	TTGTCTGTGTGTCGGGATCATTTCCCTCTGGCGGCGAGGCTGATCCGTCACGAGAGGGA	300
Query	362	ATTGAGTTTTCACAGCAAACTGATTGACGCTCTTGAAGGCGGGGTATACAGCGCTTGGGTG	421
Db	301	ATTGAGTTTTCACAGCAAACTGATTGACGCTCTTGAAGGCGGGGTATACAGCGCTTGGGTG	360
Query	422	ACTTTGTAACACTGGGATCTGCTCCTCAGGCGCTTCAAGATCGCTATGAGAGGCTGCTCAAC	481
Db	361	ACTTTGTAACACTGGGATCTGCTCCTCAGGCGCTTCAAGATCGCTATGAGAGGCTGCTCAAC	420
Query	482	GTTGGAAGAGGTCCAGCTGACCTTTGAGCGGTATGCGAGTTGTGCTTTGAACGTTTTGGG	541
Db	421	GTTGGAAGAGGTCCAGCTGACCTTTGAGCGGTATGCGAGTTGTGCTTTGAACGTTTTGGG	480
Query	542	GACCGAGTCCAGAACTGATTCACATCAACGAWCCCTGGATTACAGGCATCTATGAGAT	601
Db	481	GACCGAGTCCAGAACTGATTCACATCAACGAWCCCTGGATTACAGGCATCTATGAGAT	540
Query	602	GCCACCGGACGACGCCCCGGGACGAGACGACATTTAACAGCACTCCACGAGAGGCAAC	661
Db	541	GCCACCGGACGACGCCCCGGGACGAGACGACATTTAACAGCACTCCACGAGAGGCAAC	600

```

QY 662 ACTGCCACTGAGCGTGGCTCGCTGGAAAGGCCCAAGTCATGAGCCATGGCCGCGCTG 721
DB 601 ACTGCCACTGAGCGTGGCTCGCTGGAAAGGCCCAAGTCATGAGCCATGGCCGCGCTG 660
QY 722 GCCGTCTACAGCAGGAGCTTTCGCCCCCTCGCAAAAGGCCCAAGTCGAGCTTCGCTCAAC 781
DB 661 GCCGTCTACAGCAGGAGCTTTCGCCCCCTCGCAAAAGGCCCAAGTCGAGCTTCGCTCAAC 720
QY 782 GGGGACTACTATAGAGCCCTTGGGACAGCAATGAGCCTTCGGGACAAAGAGGCTGCTGAGCGA 841
DB 721 GGGGACTACTATAGAGCCCTTGGGACAGCAATGAGCCTTCGGGACAAAGAGGCTGCTGAGCGA 780
QY 842 CGGATGGAATTTACATTTGGCTGGTTCGCAATCCATCTTCTTGAAGAAGACTATCCA 901
DB 781 CGGATGGAATTTACATTTGGCTGGTTCGCAATCCATCTTCTTGAAGAAGACTATCCA 840
QY 902 GAGAGCATGAGAGAGCAGTGGGCGAGAGGCTTTCAGCCCTCACTCCCGGAGCTTTGCG 961
DB 841 GAGAGCATGAGAGAGCAGTGGGCGAGAGGCTTTCAGCCCTCACTCCCGGAGCTTTGCG 900
QY 962 ATCTCTAATGCCCGGAGAGACCGACTTCTAGGAGATGAAATTAACAATCCAGTTCCG 1021
DB 901 ATCTCTAATGCCCGGAGAGACCGACTTCTAGGAGATGAAATTAACAATCCAGTTCCG 960
QY 1022 CGGCACTGAGAGGCTCGCCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1081
DB 961 CGGCACTGAGAGGCTCGCCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
QY 1082 GAGAATTAAGAGCGGACCCGCTTGGCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1141
DB 1021 GAGAATTAAGAGCGGACCCGCTTGGCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
QY 1142 CGGACATGTTCCGGAAGCATCTCGCCCGGCTGTAGGCTGTAGCGGCAAGCCATCTAC 1201
DB 1081 CGGACATGTTCCGGAAGCATCTCGCCCGGCTGTAGGCTGTAGCGGCAAGCCATCTAC 1140
QY 1202 ATCAGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1261
DB 1141 ATCAGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
QY 1262 AAGGACCCCTTCGCAATCCGCTTCTTGAATCCGACTTGGACTGATTTCCAGAGCCAT 1321
DB 1201 AAGGACCCCTTCGCAATCCGCTTCTTGAATCCGACTTGGACTGATTTCCAGAGCCAT 1260
QY 1322 ACCGAGAGCGGCTCGTCTCAAGGGGATCTTTCGCGGCTGCTCGATACTTGGAA 1381
DB 1261 ACCGAGAGCGGCTCGTCTCAAGGGGATCTTTCGCGGCTGCTCGATACTTGGAA 1320
QY 1382 TGGTCAGATGGCTACGAGACCCAGATTCGCGCTACGCTCAAGACTACACACCTCAAG 1441
DB 1321 TGGTCAGATGGCTACGAGACCCAGATTCGCGCTACGCTCAAGACTACACACCTCAAG 1380
QY 1442 CGCAGCCCAAGAGTCTGCTGCTGCTCAAGAGCAATGTTTGGCGCCGCGAGAGGTT 1501
DB 1381 CGCAGCCCAAGAGTCTGCTGCTGCTCAAGAGCAATGTTTGGCGCCGCGAGAGGTT 1440
QY 1502 AAAAGTGGCGGATTA 1516
DB 1441 AAAAGTGGCGGATTA 1455

```

```

RESULT 6
US-10-026-140-3
; Sequence 3, Application US/10026140
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuurt, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697

```

```

; CURRENT APPLICATION NUMBER: US/10/026,140
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-140-3

Query Match      73.0%; Score 1454.2; DB 43; Length 1455;
Best Local Similarity: 100.0%; Pred. No. 2.2e-300;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 ATGCCGAGTCCGTACTCTGCGCCAAAGCACTTTGAATGGGGCTTGGCAACGGCGGCTAC 121
DB 1 ATGCCGAGTCCGTACTCTGCGCCAAAGCACTTTGAATGGGGCTTGGCAACGGCGGCTAC 60
QY 122 CAGATGAAAGGCGCGCTCAAGAGAGTGGCCGCGCCGCTCCATCTTGGGACAGTACTGC 181
DB 61 CAGATGAAAGGCGCGCTCAAGAGAGTGGCCGCGCCGCTCCATCTTGGGACAGTACTGC 120
QY 182 CACTGAGCATCGGCAACAAAGGCGCAACGGGATGCTTGGCTTGGATCACTACAC 241
DB 121 CACTGAGCATCGGCAACAAAGGCGCAACGGGATGCTTGGCTTGGATCACTACAC 180
QY 242 CGCTAGAGAGGACTTTGATCTTGTACCAAGTACGGGCAAAAGGCTTACCGCTTCC 301
DB 181 CGCTAGAGAGGACTTTGATCTTGTACCAAGTACGGGCAAAAGGCTTACCGCTTCC 240
QY 302 TTGTCGTGGTCGGGATCAATCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGA 361
DB 241 TTGTCGTGGTCGGGATCAATCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGA 300
QY 362 ATTAGATTTTACAGCAACTGATTTGAGCGCTGTTAGGCGGGATACAGCTTGGTG 421
DB 301 ATTAGATTTTACAGCAACTGATTTGAGCGCTGTTAGGCGGGATACAGCTTGGTG 360
QY 422 ACTTTTACACTGGGATCTGCTCGAGGCTTCAAGATTCGATAGAGCTGGCTCAAC 481
DB 361 ACTTTTACACTGGGATCTGCTCGAGGCTTCAAGATTCGATAGAGCTGGCTCAAC 420
QY 482 GTGGAAGAGTCCAGCTGAGACTTGTAGCGGATAGCAGGTTGCTTGAACCTTTGGG 541
DB 421 GTGGAAGAGTCCAGCTGAGACTTGTAGCGGATAGCAGGTTGCTTGAACCTTTGGG 480
QY 542 GACCGAGTCCAGAACTGATACACATCAAGAACCTTGAATTCAGGCACTTATGATAT 601
DB 481 GACCGAGTCCAGAACTGATACACATCAAGAACCTTGAATTCAGGCACTTATGATAT 540
QY 602 GCCACCGGAGCAAGCCCGGCGAGAGCAGATTTACAGACATCCACCGAGGGCAAC 661
DB 541 GCCACCGGAGCAAGCCCGGCGAGAGCAGATTTACAGACATCCACCGAGGGCAAC 600
QY 662 ACTGCCAGAGCGGCTGCTGCTGGAAGGCCCAATCATGAGAGCATGCGCGCGCTG 721
DB 601 ACTGCCAGAGCGGCTGCTGCTGGAAGGCCCAATCATGAGAGCATGCGCGCGCTG 660
QY 722 GCCGTCTACAGCAGGAGCTTTCGCCCCCTCGCAAAAGGCCCAAGTCGAGCTTCAC 781
DB 661 GCCGTCTACAGCAGGAGCTTTCGCCCCCTCGCAAAAGGCCCAAGTCGAGCTTCAC 720
QY 782 GGGGACTACTATAGAGCCCTTGGGACAGCAATGAGCCTTCGGGACAAAGAGGCTGCTGAGCGA 841
DB 721 GGGGACTACTATAGAGCCCTTGGGACAGCAATGAGCCTTCGGGACAAAGAGGCTGCTGAGCGA 780
QY 842 CGGATGGAATTTACATTTGGCTGGTTCGCAATCCATCTTCTTGAAGAAGACTATCCA 901
DB 781 CGGATGGAATTTACATTTGGCTGGTTCGCAATCCATCTTCTTGAAGAAGACTATCCA 840
QY 902 GAGAGCATGAGAGAGCAGTGGGCGAGAGGCTTTCAGCCCTCACTCCCGGAGCTTTGCG 961
DB 841 GAGAGCATGAGAGAGCAGTGGGCGAGAGGCTTTCAGCCCTCACTCCCGGAGCTTTGCG 900

```

```
QY 962 ATCTCAATGCGGAGAGACCGAATTCTAGGCAATGATTAACAACATCCAGTTGCG 1021
Db 901 ATCTCAATGCGGAGAGACCGAATTCTAGGCAATGATTAACAACATCCAGTTGCG 960
QY 1022 CGCCACCTAGACGGTCCCGTCCCGAGACGCACTATCTCGCGCCATCAAGACACAG 1081
Db 961 CGCCACCTAGACGGTCCCGTCCCGAGACGCACTATCTCGCGCCATCAAGACACAG 1020
QY 1082 GAGATTAAGACGGACCGCCCGTTGGGAGAGACGGCCCTCGCTGGCTGCGCTCTGC 1141
Db 1021 GAGATTAAGACGGACCGCCCGTTGGGAGAGACGGCCCTCGCTGGCTGCGCTCTGC 1080
QY 1142 CGGAGACGTTTCGGAGACATCTGCGCCGGGTGACGCGCTGTAAGGCAAGCCCATCTAC 1201
Db 1081 CGGAGACGTTTCGGAGACATCTGCGCCGGGTGACGCGCTGTAAGGCAAGCCCATCTAC 1140
QY 1202 ATCAACGAGACGGATCCCGTCCCTGAGAGAGAGAACTGACGTGCGAGAGCCGTC 1261
Db 1141 ATCAACGAGACGGATCCCGTCCCTGAGAGAGAGAACTGACGTGCGAGAGCCGTC 1200
QY 1262 AAGCAACCTTCCGCAATCCGTAATTTGACTCGCACTTGACCTGATTTCCAGGCCATT 1321
Db 1201 AAGCAACCTTCCGCAATCCGTAATTTGACTCGCACTTGACCTGATTTCCAGGCCATT 1260
QY 1322 ACCAGAGACGGCGTCTGTCAGAGGGGTACTTTGCGGGCGTGTGCTGATACTTGAAA 1381
Db 1261 ACCAGAGACGGCGTCTGTCAGAGGGGTACTTTGCGGGCGTGTGCTGATACTTGAAA 1320
QY 1382 TGGTCAGATGCTACGAGACCCAGATTGCGCGTCAAGTTCAACACCACTTCAAG 1441
Db 1321 TGGTCAGATGCTACGAGACCCAGATTGCGCGTCAAGTTCAACACCACTTCAAG 1380
QY 1442 CGCAGCGCCCAAGAGTCTGCGCTTGTCTCAAGACATGTTTGGGCGCCGCAAGGTT 1501
Db 1381 CGCAGCGCCCAAGAGTCTGCGCTTGTCTCAAGACATGTTTGGGCGCCGCAAGGTT 1440
QY 1502 AAAGTGGCGGCACTAA 1516
Db 1441 AAAGTGGCGGCACTAA 1455
```

```
RESULT 7
US-09-533-559-7458
; Sequence 7458, Application US/09533559
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjarke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/09/533.559
; EARLIER FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 09/273.623
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Tricoderma reesei
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(735)
; OTHER INFORMATION: n = A,T,C or G
US-09-533-559-7458
```

Query Match 30.0%; Score 598; DB 22; Length 735;
Best Local Similarity 97.0%; Pred. No. 2.9e-117;

```
Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;
QY 319 CATTCCTCTGGGAGAGGCTGATCCCGTCAACGAGAGGAATGATTTACAGCAA 378
Db 1 CATTCCTCTGGGAGAGGCTGATCCCGTCAACGAGAGGAATGATTTACAGCAA 60
QY 379 ACTGATTTGACGCCCTGTTGAGGCGGGGTATCAAGCCTTGGGTGACTTTGTACACATGGGA 438
Db 61 ACTGATTTGAGGCCCTGTTGAGGCGGGGTATCAAGCCTTGGGTGACTTTGTACACATGGGA 120
QY 439 TCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTCAAGTGAAGAAGGTCAGGT 498
Db 121 TCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTCAAGTGAAGAAGGTCAGGT 180
QY 499 GGACTTTGAGCGGTATGCGAGGTTGTGCTTTGAAAGTTTTGGGACCGAGTCCAGAACTG 558
Db 181 GGACTTTGAGCGGTATGCGAGGTTGTGCTTTGAAAGTTTTGGGACCGAGTCCAGAACTG 240
QY 559 GATCACCATCAACGAGCCCTGATTTCAAGGCTATGATGATGATGCAACCGGACAGCAAGC 618
Db 241 GATCACCATCAACGAGCCCTGATTTCAAGGCTATGATGATGATGCAACCGGACAGCAAGC 300
QY 619 CCGGAGAGAGACGATTTAAACAGCACTTCAACGAGGAGAACTGAGCCGTCG 678
Db 301 CCGGAGAGAGACGATTTAAACAGCACTTCAACGAGGAGAACTGAGCCGTCG 360
QY 679 GCTCGCTGAAAGGCCCAAGATCATGAGCCATGCGCGCGCGTGTCTTACAGCAGAGGA 738
Db 361 GCTCGCTGAAAGGCCCAAGATCATGAGCCATGCGCGCGCGTGTCTTACAGCAGAGGA 420
QY 739 CTTTGGCCCTCGCAAAAGGCGCAGATGCGCATCTGCTCAAGCGGCACTATGAGCC 798
Db 421 CTTTGGCCCTCGCAAAAGGCGCAGATGCGCATCTGCTCAAGCGGCACTATGAGCC 480
QY 799 CTGGAACAGAAATGAGCTCTGGAACAGAGGCTGCTGACGACGAGTGAATTTCAAT 858
Db 481 CTGGAACAGAAATGAGCTCTGGAACAGAGGCTGCTGACGACGAGTGAATTTCAAT 540
QY 859 TGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTA-TCCAGAGGATGAAGAGC 917
Db 541 TGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTA-TCCAGAGGATGAAGAGC 600
QY 918 AGCT--GGGCGAGAGGCTTCAGGCTTAC-TCCCGGCACTT--GCCATCTCAATGC 972
Db 601 AACTTGGGCGAGAGGCTTTCACACTTCTCCGCGGACTTTTGCATCTTNAATGC 660
QY 973 CGGAGAGACC 982
Db 661 CGGAGAGAAC 670
```

```
RESULT 8
US-10-653-047-7458
; Sequence 7458, Application US/10653047
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjarke Olsen
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/10/653.047
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US/09/533.559
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/273.623
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7458
```

```
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Tricoderma reesei
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(735)
; OTHER INFORMATION: n = A,T,C or G
US-10-653-047-7458
```

Query Match 30.0%; Score 598; DB 53; Length 735;

Best Local Similarity 97.0%; Pred. No. 2,9e-117; Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;

```
QY 319 CATTCCTCGCGCGAGCTGATCCCTCAACGAGAGGAAATTGATTTACAGCA 378
Db 1 CATTCCTCGCGCGAGCTGATCCCTCAACGAGAGGAAATTGATTTACAGCA 60
QY 379 ACTGATTGACGCCCTGTTGAGCGGGGATACGCCCTTGGGTGACTTTGTA 438
Db 61 ACTGATTGACGCCCTGTTGAGCGGGGATACGCCCTTGGGTGACTTTGTA 120
QY 439 TCTGCTCAGGCGCTTACGATGCTATGAGAGCTGGCTCAAGTGAAGAGTCCAG 498
Db 121 TCTGCTCAGGCGCTTACGATGCTATGAGAGCTGGCTCAAGTGAAGAGTCCAG 180
QY 499 GCACTTTGAGCGGATGATGAGGTTGCTTTGAAAGTTTGGGAGACGAGTCCAG 558
Db 181 GCACTTTGAGCGGATGATGAGGTTGCTTTGAAAGTTTGGGAGACGAGTCCAG 240
QY 559 GATCACCATTCAACGAGCCCTGATTCAGGCACTCTATGATGATGCCACCGG 618
Db 241 GATCACCATTCAACGAGCCCTGATTCAGGCACTCTATGATGATGCCACCGG 300
QY 619 CCGGCGAGAGAGCACTTAAACAAGCACTCAACGAGGCAACACTGCCACTGAG 678
Db 301 CCGGCGAGAGAGCACTTAAACAAGCACTCAACGAGGCAACACTGCCACTGAG 360
QY 679 GCTCGTGAAGAGCCAGATCATGAGCAATGCCCGCGCGCTGCGCTTACAGCA 738
Db 361 GCTCGTGAAGAGCCAGATCATGAGCAATGCCCGCGCGCTGCGCTTACAGCA 420
QY 739 CTTTCGCCCTCGCAAAAGGCGCAGATGCGCATCTCGCTCAACGCGCACTA 798
Db 421 CTTTCGCCCTCGCAAAAGGCGCAGATGCGCATCTCGCTCAACGCGCACTA 480
QY 799 CTGGGACAGAAATGAGCTTGGGCAACAAGAGGCTGCTGAGCGAGATGGA 838
Db 481 CTGGGACAGAAATGAGCTTGGGCAACAAGAGGCTGCTGAGCGAGATGGA 540
QY 859 TGGCTGATTTGCAATCCCATCTTCTTGAAGAAGACTA-TCCAGAGAGTGA 917
Db 541 TGGCTGATTTGCAATCCCATCTTCTTGAAGAAGACTA-TCCAGAGAGTGA 600
QY 918 AGCT--GGGCGAGAGCTTCCAGCCCTCAC-TCCCGCGAGCTTT--GCAT 972
Db 601 AGCTGGGGGAGAGGCTTTCAACCCCTTACTTCCCGGAGCTTTTGCATCT 660
QY 973 CGGAGAGACC 982
Db 661 CGGAGAGAAC 670
```

RESULT 9

```
US-09-417-507-9340
; Sequence 9340, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; TITLE OF INVENTION: FUNGICIDES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417, 507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
```

```
; SEQ ID NO 9340
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: A.fumigatus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1562)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-417-507-9340
```

Query Match 17.3%; Score 345.4; DB 20; Length 1572;

Best Local Similarity 55.7%; Pred. No. 3,8e-63; Matches 721; Conservative 2; Mismatches 563; Indels 9; Gaps 3;

```
QY 88 CGACTTTGATGAGGCTTGCAACGCGCGCTTCCAGATGAGAGGCGCGCTCAAG 147
Db 66 CGACTTTTTCATGATGTTACCCACCGGGGCGCCAGGTGAGAGACTTGGATTAA 125
QY 148 TGGCCGCGCGCGCTTCCATCTGGGACAGTACTGCACTGGAGCCATCGGCA 207
Db 126 CGGTAAAGGCCATCAATCTGGGATCTTTTGGGACACACCTGGGAAAGTGA 185
QY 208 CGCCAAACGCGAGATGCTTGGCATCTACACCGCTACAGATGAGACTTTGAT 267
Db 186 AACCACTGGGAGAGAGCGCGGTACGCTGTATGACCTGTACAGAGATGCG 245
QY 268 GACCAAGTACGGGCAAAAGGCTTACCGCTTCTGCTGCTGCGGATCAATCC 327
Db 246 GAAGTGTATGGGCTCAATGATACAGCTTCTCTCTGCTGCTGCTGCTGCT 305
QY 328 CCGCGCGAGCTGATCCCTTCAACGAGAGAGGAAATTGATTTACAGCAACT 387
Db 306 TGGAGGCTGCGAGAGCTGCTGCAACGAAAGGAGATTGAACTATTCAATCT 365
QY 388 CGCCTGTTGAGCGGGGATACAGCCTTGGGATCTTGTACCACTGGATCTG 447
Db 366 CGAGCTTCTCCGTAATGAGCATACACCGCTTGTGCTTCTTCCACTGGG 425
QY 448 GGGCTTCAAGATGCTATGAGAGGCTGCTCAACGTGGAAGAGTCCAGTGA 507
Db 426 ATGCTGAGAGATGCTACGCGCGCATGCTGAACAGAAATTTGTGCGGCT 485
QY 508 GCGGTATGCGAGGTTGTGCTTTGAACGTTTGGGAGCCGAGTCCAGAACT 567
Db 486 CAATTACGCTCGGTTTGTCTTGAAGAGACTCGGCGATTCGCTGAAGATT 545
QY 568 CAACGAGCTGATTCAGGCTTCTATGAGATATGCCACCGGAGCAACGCG 627
Db 546 CAACGAGCTGATTCAGGCTTCTATGAGATATGCCACCGGAGCTCACCG 605
QY 628 GAGCAGATTAAACAAGCATCCACCGAGGAGCAACTGCCACTGAGCCGCT 687
Db 606 GTCTTATTCGCGGACCGCAACGAGAGGCGGCTGCTCACCGAGCATTC 665
QY 688 AAGAGCCAGATCATGAGCCATCCGCGCGTGGCGCTTACAGAGAGGACT 747
Db 666 GCAACGAGAGCTGTGCGCCACGAGGACGCTGCTTCTTACAGAGAGTTC 725
QY 748 CTGCAAAAGGCGCAATCGCATCTTCTCAACGCGGACTACTATGAGCC 807
Db 726 CCATTCAGCAAGGAGCATCGGCATTCACGCTTACGGAACCTGCTCG 785
QY 808 CAATGAGCTTCGAGCAAGAGAGCTGAGCGAGCGAGATGGAATTTCAT 867
Db 786 AGCGAGCCGCTGATCAAGCGCGCGCTGAGCGCGCGGAGATTCAGAT 845
QY 868 TGGCAATCCCATCTTCTTGAAGAAGACTACAGAGAGCATGAAGAGCT 927
Db 846 CCGGAGCCGCTTACAGAGCGGATGATCTCGGCTGATGAGAGCCAGCT 905
QY 928 GAGGCTTCAACGCTCACTCCGCGGACTTTGCAATCTTCATGCGGAGAG 987
Db 906 CCGCTGCTCCAAAGTACAGCCCGAGAGAT---CAAAGCTGTGCTTGGAG 962
```

QY 988 CTACGCGATGAATTAATTAACAATCCAGATTGCGCGCCCACTAGACGGTCCCTCCCGA 1047
DB 963 CTATGGCATGAATCTGTACACAGACTTCTTCTGTCAACACAGACACCGCGGAGCAT 1022
QY 1048 GACGAGTATATCTGCGCCCATCAATGAGCACAGAAATATAGAGCGGCGCCGTTGG 1107
DB 1023 TAAACAGACACAAAGGGAATGAGATTCATGATTTCAACAAACATGGCATCTCGAGG 1082
QY 1108 CGAGAGAGCGGCTCGCTGCTGCGCTGCTGCGCGGACATGTTCCGAGACATCTCGC 1167
DB 1083 CGAAGAGAGCGACACCGAGTGTCTGCAAGCTGCGCCCTCGGAGGCTTCGAA 1142
QY 1168 CCGGATGTACGCTGTATCGGCAAGCCCATCTACATCAACGAGACGATCCGTGCCC 1227
DB 1143 TTGATCTGTCTGTGTACAGATGCCCATCTATGTACGAGAAATG--GCACGACAGC 1199
QY 1228 TGAAGAGAAACATGATGTCGAGAGAGCGGCTCAACGACCCCTCGCATCCGTAATT 1287
DB 1200 CAAGGCGAGACAGCCCTTCAACCCAGTGTCTCAACGACCAAGTTTGCATCAGGTTCT 1259
QY 1288 TGAATCGCATTTGACT--CGATTTCCAGGCGCATTAACGAGACGCGGTGTCTCA 1344
DB 1260 TGAAGGATATGTGGGTGGGCACTGGGCTGTGCGGTCAAGAGACGGGATTGATATCG 1319
QY 1345 GGGGACTTTGGGTGGGCGTGTGCTGATTAATTGG 1379
DB 1320 CTCCATTTTGTGCTGACATTCACGATTAACCTGG 1354

RESULT 10

US-09-417-507-19268
; Sequence 19268, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; FILE REFERENCE: PAT99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 19268
; LENGTH: 741
; TYPE: DNA
; ORGANISM: A.fumigatus
US-09-417-507-19268

Query Match 16.8%; Score 334.4; DB 20; Length 741;
Best Local Similarity 67.3%; Pred. No. 7.3e-61;
Matches 470; Conservative 1; Mismatches 227; Indels 0; Gaps 0;
QY 682 CGGTGAAAGGCCAGATATGAGCCAGCCGCGCGCTGCTTACAGAGGACTT 741
DB 6 CGCTTGTCTGCGCCATCCCGGTGTGAATACCGCGCGCTGCTGTACACAGAGATT 65
QY 742 TGCGCCCTCGCAAAAGGCGCAATCGCATCTGCTCAACGCGCACTAATAGCCCTG 801
DB 66 CGGATGTTCACAGGCAAGATAGGATCTGTTGAAGAGACTATAGAGCTTG 125
QY 802 GGAACAGATGAGCTCTGAGCAAGAGGCTGTGAGCAAGAGTGAATTTTCACTTGG 861
DB 126 GAATGCCAGAAAGAGCGGATCAGCTGTGCAAGAGCGTGCATGAATTTTCACTTGG 185
QY 862 CTGATTTGCAATCCATCTTCTTGAAGAGACTTCAAGAGAGATGAAGAGAGCT 921
DB 186 CTGATTTGCAATCTTCTTCTTGAAGAGACTTCAAGAGAGATGAAGAGAGACT 245
QY 922 GGGCAGAGAGCTTCAGAGCTCACTCCGCGGACTTTGCGCATCTCAATGCGGAGAGC 981
DB 246 TGAAGCTCGCTTACAGAGTTTCCCTCTGATTTTGAAGCTCTTCCGAGGCGAGAG 305
QY 982 CGACTTCTACGAGATTAATTAACATCCAGTTTGCGGCGCACTTGAAGAGTCCCT 1041

DB 306 CGATTTCTACGGGATGAATTAATTAATCCAGTTTGTCTGCGCATTCAGTCAACAGC 365
QY 1042 CCCCCGAGAGCATATCTTCTGCGCCATCCATGAGACAGAGAAATTAAGACGAGGCC 1101
DB 366 ATCCGAAACAGATTAATTAATGAGCAATGATGATTAACAGAGAAACATTAAGGAGACTC 425
QY 1102 CATTGCGAGAGAGCGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1161
DB 426 AGTTGGCAACCAAGTGTATCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 485
QY 1162 TCTGCGCGGCTGTACGCTGTATCGGCAAGCCCATCTCAATCAACCGAAGAGAGTCC 1221
DB 486 CTTCAAGAGATTAATGCTGTATCGCTGTATCGGCAACCAATTTTCAATCAGAGAGAGTCC 545
QY 1222 GTGCGCTGAGAGAGAGAAATGACGTGCGAGAGAGCGCTCAACGACCCCTTCGATCCG 1281
DB 546 ATGTCTGAGAGAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605
QY 1282 STACTTGAATCTGCACTTGTGATGATTTTCAAGGCAATTAACGAGAGGCGTGTCT 1341
DB 606 GTATTTTGAAGATCATCTAGAGGCTGTGCGCTGTGTCAACCAAGAGGCTGTATAT 665
QY 1342 CAAGGCTACTTTCGCTGCGGCTTGTGCTGATTAATTGG 1379
DB 666 CAGAGGCTACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 703

RESULT 11

US-09-404-520-15403
; Sequence 15403, Application US/09404520
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Ghodssi, Azila
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: McIninch, James
; APPLICANT: Timberlake, William B.
; TITLE OF INVENTION: Emericella nidulans Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15498)A
; CURRENT APPLICATION NUMBER: US/09/404,520
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 44345
; SEQ ID NO 15403
; LENGTH: 2918
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-09-404-520-15403

Query Match 16.0%; Score 318.6; DB 20; Length 2918;
Best Local Similarity 53.8%; Pred. No. 2.4e-57;
Matches 701; Conservative 1; Mismatches 595; Indels 6; Gaps 2;
QY 77 GCTCTGCCAAGCACTTGTGAATGAGGCTTCCGCAACGCGCGCTTACAGATGAAAGCGCC 136
DB 1147 GCTTGGCAATGACTTCTTCCATGATATGCAACGCGCGGCGGAGTTGAAGAGCG 1206
QY 137 GTCAAGAGGTGCGCGCGCGCTGCTGATCTGAGCAAGTACTGCCACCTGGAACCATCG 196
DB 1207 TGAACAGAGATGGAAGGCTCTTGCATCTGGGACATTCGCGCATCTCCAGGCAAA 1266
QY 197 CGCACCAAGCGCGCAACGCGATGTGGCTTGCATCACTACACCGCTACGATGAGAC 256
DB 1267 GTCAAGCAATTAAGCAATGCGATGACGCTGTGAGTTTCAAGCTTTTACCGGAGAT 1326
QY 257 TTTGATCTCTTGAACCAATGAGCGGCAAGGCTTCTCTCTTGTGCTGTGCGG 316
DB 1327 GTTGGCCGATGAGATGTATGATGATCAACCTTATCATCTGCTATCTCTGTGCGG 1386
QY 317 ATCATTTCCCTGCGCGGAGGCTGATCCCTGCAACGAGAGGAAATGAGTTTACAGC 376
DB 1387 ATTAATCCCACTCGAGGCGCGGACGACCGGTCAATGAGCAAGGATTAAGTACTACAG 1446
QY 377 AATAGTATGACGCGCTGTGAGCGGAGTATCAGCGCTTGGGTGATCTTGTATCACTGG 436

```

Db      1447 GACCTGGTAGACGAGCTACTCAATATGAGATTACGCCATTGTAACCTCTTCACTGG 1506
Qy      437 GATGTGCTGAGGGCTTCAAGATGCTATGAGAGGCTGCTCAACGAGGAAGGTCCAG 496
Db      1507 GATGTGCTGAGGGCTTCAAGATGCTATGAGAGGCTGCTCAACGAGGAAGGTTCATT 1566
Qy      497 CTGAGCTTTGAGCGGTATGCGAGGTTTGCTTTGAAGCTTTTGGGAGCCAGTCCAGAAC 556
Db      1567 CTGAGCTTTTCCGCTACGCGCGTGTCTGTTTGAAGCCTTAGGCCAAAGTCCGTCAC 1626
Qy      557 TGGATCAACATCAACGAGCCCTGAGTTCAGGCCATCTATGATATGCGACCGGAGCAAC 616
Db      1627 TGGATTAACATTAATGAACCTGGGTGTACTCGCTGACAGGTATGCGGAGGTTTCAT 1686
Qy      617 GCGCCGGGAGGAGGAGCAGATTAACAAGCATCCACCGAGGCAACATGCGCATGAGCCG 676
Db      1687 GCTCTGTCAGCGGTGATGTTTCCGAACTCATGTAAGAGGATGATTCCTGACAGAACCG 1746
Qy      677 TGGCTGCTGGAAGGCCAGATCATGAGCCATGCGCGCCGTGCGCTTACAGCAGG 736
Db      1747 TTTATGCTGCGCCACACAAATGCTGTACCCAGGCCATGTTTCCAAAGCTTTACCGAG 1806
Qy      737 GACTTTCGCCCTTCGCAAAAGGCGCAGATCGGACTCTGCTCAACGCGCATACTATGAG 796
Db      1807 GTGTTCCAAACCGAGCAGAAAGAAACAATTGGCATCACTCCATGCAACTGCTGCGAA 1866
Qy      797 CCTGTGAGCAGCATGAGCTTCGGGACAAAGAGGCTCTGAGCAGCAGGATGGAATTTAC 856
Db      1867 CTTTGGAGTAAAGACGATTCGCGGAGCAGAGAGCAGAACAGAGCGGCAAGGGAATTTGAG 1926
Qy      857 ATTGGCTGTTTCCAAATCCCATCTTCTTGAAGAAGGACTATCCAGAGAGCATGAAGAAG 916
Db      1927 ATGCTGTGTTTCCCGGACCCCTTATACAAAGCAGGCGCATACCCAGGCTCAATGCGGCT 1986
Qy      917 CAGCTGGGAGGAGGCTTTCAGGCTTCAGCTCCGCGGAGCTTTGCGCATCTCAATGCGGAG 976
Db      1987 CAATCGGGAGCGGTCTTACCGCGTTTCACTCCGAGAGGT---CAAAATGCTACTGAGA 2043
Qy      977 GAGACCGACTTCTACGCGCATGAATTAACAATCCAGATTCGCGGCGCATTAAGACGCT 1036
Db      2044 AGTTCAAGAAATTTTACGCTATGAATCGTACACGACTTTTTCGTCAGCAGCAAGGATACG 2103
Qy      1037 CCGCTCCCGGAGCGGACTATCTCGGCGCCATCCATGAGCAGCAGGAATTAAGAGCGG 1096
Db      2104 CTTCCAGATATCAATACCAACAAAGGCAATGCTATTCTTACGATACAAACAGAAAGGC 2163
Qy      1097 AGCCCGCTGGGAGGAGAGCGGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1156
Db      2164 GTCTTCGAGGCGGAAGGTCCGATAGCGCTTGGCTGCGGAGCGGACCCACTGGATGCGC 2223
Qy      1157 AAGCATCTCGCCCGGGGTGACGGCTGTACGGCAAGCCCATCTACATCAGCAGAGAGCGA 1216
Db      2224 AAGCTCTCAATTTGATTTGGAACCGATATCATGTGCTTATATATGTCAGGAAAGT-- 2281
Qy      1217 TGCCTGCGCTGAGAGAGAGAAATGACGTCGAGAGAGGCGCGCAACGACCCCTTCCGC 1276
Db      2282 -GCACAGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy      1277 ATCCGATCTTTGATCTGACCTTGAATGCTGATTTTCCAAAGCCATTAACAGAGAGCGCTC 1336
Db      2341 ATGCGATTTCTTGAAGGCTAGTGTGATGATTTGGCCCGTGTCTGTCAAAAGAGAGATGAGTG 2400
Qy      1337 GTGCTCAAGGGGATCTTTGCTGGGCTGTGCTGATTAACCTGG 1379
Db      2401 GATATCCGATCTTAATTTGCTGGAACCTTCAACGAGCAATGCGG 2443

```

RESULT 12
 US-10-369-493-37092
 ; Sequence 37092; Application US/10369493
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei

```

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 37092
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-37092

Query Match      16.0%; Score 317.8; DB 50; Length 1470;
Best Local Similarity 53.8%; Pred. No. 3e-57;
Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;

Qy      81 TGCCTCAACGACTTTGAATGAGGCTTGGCAACGCGCCCTACAGATGGAAGCGCCCTCA 140
Db      2 TGGCGAATGACTTCTTCATGATATGCAACGCGCGCGGAGGTTGAAGAGCTGGA 61
Qy      141 AAGAGGTGGCGCGCGCGCTTCATGAGGACAGTACGTCGCACTGGAAGCATCGGCA 200
Db      62 ACAAGATGGCAAGGCTCTTCATGAGGACACATTCGGGCACTCCAGGCAAGTCA 121
Qy      201 CCAAGCGCGCAACGCGGATGTGCTTGCATCACTCAACCGCTACAGATGAGAGACTTTG 260
Db      122 AAGACATATACATATGGGATGAGCTGTGAGTTCTACAGCTTTTACCGGAAGATTTG 181
Qy      261 ATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTTGTGCTGCTGCGATCA 320
Db      182 CCGTGAATGAATGTATGTGTCAACGCTTATGATTCGCTATCTGCTGCGGCAATA 241
Qy      321 TTTCCCTCGGCGGAGGCTGATCCGCTCAACGAGAGGGAATTTGATTTACAGCAAC 380
Db      242 TCCACTCGGAGGCGCGGACGACCGGTCATAGCAAGGGAATCAATACTACAGAGAC 301
Qy      381 TGAATGAGCCCTGTTGAGGCGGAGGTATCAAGCTTGGTGTGACTTTGTACACTGGATC 440
Db      302 TGGTACAGAGCTACTCAATATGAGATTAAGGCTATTCGTAACCTTTTCACTGGATG 361
Qy      441 TGCCTAGGCGCTTACGATGCTATGAGGCTGCTCAACGTTGAAGAGGTCAGCTGG 500
Db      362 TTTCCGAGGCTTTGAGAGATCGGTACGCGGCGATGTGAACAGAGAGGTTTACTCTG 421
Qy      501 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGAGCCAGTCCAGAACTTGA 560
Db      422 ACTTTGCGGCTACGCGCGGTGTGTTTGAAGCGCTTGAAGGCAAGAGTCTGTCACTGA 481
Qy      561 TCACATCAACGAGMCCCTGATTCAGGCACTATGATATGCAACGCGAGCAACGCC 620
Db      482 TTACATTTCAATGAACCTGTGTGTACTCGCTGGCAGGTTATGCGGAGGTTTCACTGTC 541
Qy      621 CGGCGAGAGACAGCATTAACAGCACTCCACGAGGCAACACTGCCATGAGCGGTGGC 680
Db      542 CTGACGCTCATGCTTTTCGGAATCAATGAAAGGATGATTCCTGACAGAACCGTTCA 601
Qy      681 TGCCTGAAGAGCGCCAGATATAGACATAGCCCGCGCGGTGCGCTTACAGAGAGACT 740
Db      602 TCGTGGCCCAACAAATCTGTGACCCACGCGCATGTTTCAAGCTTTTACAGAGAGTGT 661
Qy      741 TTTGCCCTTCGCAAAAGGCGCAGATCGGCACTCTGCTCAACGCGGCACTATATAGCCCT 800
Db      662 TCCACCGCAGAGAAAGAGAAACAATTGATATCACTCTCATGCACTGTCGGAACCTT 721
Qy      801 GGGACAGCAATGAGCTTCGGGACAGAGGCTGCTACGAGCAGATGAATTTCACTTG 860
Db      722 GGGATGAAGACGATTCGCGGAGCAAGAGAGCAACAGGAGCAGGGAATTCGAGATCG 781

```


QY 1341 TCAAGGGTACTTTCGTCGGGCGTTCGATTAATTGG 1379
 Db 1256 TCCGGTCTTACTTTCGTCGGGCGTTCGATTAATTGG 1294

RESULT 14
 US-60-138-103-7477
 ; Sequence 7477, Application US/60138103
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Timberlake, William E.
 ; TITLE OF INVENTION: ASPERGILLUS NIDULANS GENOME SEQUENCE AND USES THEREOF
 ; FILE REFERENCE: 38-10(15485)D
 ; CURRENT APPLICATION NUMBER: US/60/138,103
 ; NUMBER OF SEQ ID NOS: 28006
 ; SEQ ID NO 7477
 ; LENGTH: 2914
 ; TYPE: DNA
 ; ORGANISM: Aspergillus nidulans
 US-60-138-103-7477

Query Match 15.4%; Score 307.6; DB 69; Length 2914;
 Best Local Similarity 53.8%; Pred. No. 5,4e-55;
 Matches 701; Conservative 1; Mismatches 595; Indels 7; Gaps 3;

QY 77 GCTCTGCCCAACGACTTTGAATGGGCTTCGCAACGGCCGCTTACCGATGCAAGGCGC 136
 Db 1147 GCCTTGGCAATATCTTTCATGATGATGCGCCGCGCGCGCGGCTTGAAGGCG 1206
 QY 137 GTCAAGAAAGTGGCGCGCGCGCGCTTCATGCGGACGATGCTGACCTGAGCCATCG 196
 Db 1207 TGGAAACAGATGGCAAGGCTCTTCGATCTGGACATTCGGGCACTCCAGGCAAA 1266
 QY 197 CGCAACCAACGCGCAACGCGGATGTCCTTCGATCACTACCAACGCTTACGATGAGAC 256
 Db 1267 GTCAAGAAACATAGCAATGCGGATGACGCTGTAGGCTTACGACTTTTACCGGAAGAT 1326
 QY 257 TTTGATCTTTCGCAACGATGCGGCGCAAGGCTTACCGCTTCTTGTGTGTCGCGG 316
 Db 1327 GTTGCCCTGATGAAGTGTATGTCATGTCACGCTTATGATCTTCGCTATCTGCTCGC 1386
 QY 317 ATCATTCCTCCGCGCGGCGGATCCGCTCAACGAGGAGGATTTAGTTTACAGC 376
 Db 1387 ATATATCCACTCGAGGCGCGCGGACGACCCGCTCATATAGCAAGGATCAAGTACTACAG 1446
 QY 377 AAATGATTCAGCCCTCTTTCGAGCGGCGGATACGCGCTTGGTGAATTTGACCACTGG 436
 Db 1447 GACTGTGTAGACGAGCTATCACTAATATGGAATTAAGGCTTTCGTAACCTCTTCACTGG 1506
 QY 437 GATTCGCTCAGGCGCTTACGATGCTATGAGAGGCTGCTCAACGTTGAAGAGTCCAG 496
 Db 1507 GATGTTCGCGAGGCTTTCGAGGATCGGTACGCGGCGCATGTCTGAACGAGAGGTTTATT 1566
 QY 497 CTGAGCTTTCAGGCGTATGAGGTTGTCCTTGAAGCTTTTGGGAGCCAGTCCAGAAC 556
 Db 1567 CTGACTTTCGCTACGCGCGCTCTGTTTCAGAGCCCTAGGAGCCAAAGTCCGTAC 1626
 QY 557 TGGATCACTCAACGAGMCCCTGG-ATTGAGCCATCTATGATATGCAACCGGACAGAA 615
 Db 1627 TGGATTCATTCATTAAGAAATGATGTGTACTGCTGAGGAGTTATCGGAGGTTTCA 1686
 QY 616 CGCCCGCGGAGAGGAGCATTAACAGCACTCCACGAGGCGCAACTGCACTGAGCC 675
 Db 1687 TGGCTCCGACGCGCTATGTTTCGGAATCAATGAAGAGGTTATTCGACAGAAC 1746
 QY 676 GTGGCTGCTGGAAGGCGGATGATGAGGATGCGCGCGCTGCGCTTACAGACAG 735
 Db 1747 GTTCATGTCGCGCACCAAAATGTTGACCAAGGCGCATGTTTCAAGCTTTACAGAGA 1806
 QY 736 GGAATTCGCGCTTCGCAAAAGGCGCATGTCGATCTGCTCAACGCGCATCTATGA 795

Db 1807 GGTGTTCCAAACGACAGAAAGAAACAATTGGCATACCTTCATGCGCACTGTGCGA 1866
 QY 796 GCCCTGGAGACAGATAGAGCTTCGGAGCAAGAGGCTGTGAGCGACGATGAAATTTCA 855
 Db 1867 ACCTTGGATGAAGAGATTCGGGAGCAAGAAAGACAGACCGGCGGCAAGATTCCA 1926
 QY 856 CATTCGCTGTTTCCAAATCCATCTTTCGAAAGAACTATCCAGAGATGAAGAA 915
 Db 1927 GATCGCTTGGTTCCCGAGCCCTTATCAAGACAGGACTACCCAGCTCAATGCGGCG 1986
 QY 916 GCAGCTGGGCGAGGCTTCAGCCCTCACTCCCGGAGCTTGGCATCTCAATCCCG 975
 Db 1987 TCAATCTGGGAGCGCTTACCGGTTTCACTCCGAGAGT---CAAAATGTTACTAGG 2043
 QY 976 AGAGACGACTTTCAGGCGATGATTAACATCCAGTTTCGCGCGCCACTGAGCG 1035
 Db 2044 AAGTTCAAGATTTTACGATGATGACTGTCACGACTTTTTCGACAGCAAGATAC 2103
 QY 1036 TCCCGTCCCGAGACGAGATCTCGCGCCATTCATGAGCACAGAGATTAAGACGG 1095
 Db 2104 GCCTCCAGATATCAATGACCAAAAGGCAATGTCATGTTTCAAGATCAAAAGCAAGG 2163
 QY 1096 CAGCCCGTTTGGGAGAGAGCGGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1155
 Db 2164 CGTCTCTGAGGCGAAGAGTCCATACGCTTGTGCTGCGGACGCACTGATGCGG 2223
 QY 1156 GAGCATCTTCGCGCGGCTGATAGCGGCAAGCCATCATACCAAGCAAGAAAGG 1215
 Db 2224 CAGGCTGCTCAATTTGATTTGAAACGATATCATGTCCTTATATGTCACGAAATG- 2282
 QY 1216 ATGCGCGTCCCTGAGAGAGAGAAATGACGTGCGAGAGCGCGTAAACGACCCCTTCG 1275
 Db 2283 --GCAGAGACGAAAGAGAGAGACAGCTCCACACAGAGGTAATCATGACACTTTCG 2340
 QY 1276 CATCCGTAATTCGATCCGACTTTCGATGATGATTTTCAAGGCGATTAACAGAGCGCT 1335
 Db 2341 TATGCAATCTTCGAAAGGCTACCTTGTGATGATGCGCGCTGCTGTCAAGAGATGAGT 2400
 QY 1336 CGTCTCAAGGCGTACTTTCGCGGCGTTCGATGATTAATTGG 1379
 Db 2401 GATATCCGCTTACTTTCGCTGACCTTACCGCAACTGG 2444

RESULT 15
 US-09-404-520-19929
 ; Sequence 19929, Application US/09404520
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Ghodse, Azila
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: McIninch, James
 ; APPLICANT: Timberlake, William E.
 ; TITLE OF INVENTION: Emericella nidulans Genome Sequence and Uses thereof
 ; FILE REFERENCE: 38-10(15498)A
 ; CURRENT APPLICATION NUMBER: US/09/404,520
 ; NUMBER OF SEQ ID NOS: 44345
 ; SEQ ID NO 19929
 ; LENGTH: 1434
 ; TYPE: DNA
 ; ORGANISM: Aspergillus nidulans
 US-09-404-520-19929

Query Match 14.8%; Score 294.6; DB 20; Length 1434;
 Best Local Similarity 53.7%; Pred. No. 2.8e-52;
 Matches 676; Conservative 1; Mismatches 575; Indels 7; Gaps 3;

QY 122 CAGATGAAGGCGCGCTCAAGAAAGTGGCGCGCGCGCTTCATCTGCGGACAGTACTGC 181
 Db 3 CAGGTGAAGAGCGGTGAACAAAGATGGCAAGGCTCTTCGATCTGCGGACATTCGGG 62
 QY 182 CACTGAGCATTCGCGACCAACGCGGCAAGGCGGATGTCGATCACTACAC 241

[illegible]

Db	63	CATACCTCAGGGAAGATCAAGAACAATAGCAATGGGATGAGCGTGTAGAGTTCTACAC	122
OY	242	CGCTACGATGAGGACTTTGATCTCTTGACCAAGTACGGCGCAAGGCGCTACCGTTCTCC	301
Db	123	TTTTAACCCGGAAGATGTTGCCCTTGATGAAAGTGTATGTGTGTCACACGCTTAATCGATTCTCG	182
OY	302	TTGTGTGTGTGGCGGATCATTCCTCCGGCGCGCAGCTGATACCCGTCAAGAGAGGGA	361
Db	183	CTATCTGTGTGGCGATATATCCCACTCGGAGGCGCCGACGACCGGTTCAATGAGCAAGG	242
OY	362	ATTGAGTTTATCAGCAAACTGATTGACCGCCGTGTGAGGCGGGGATACACGCTTGAGGTG	421
Db	243	ATCAAGTACTACAGAGACTGGTATACAGAGCTACTCAATATGGATTATCCCATTTCCGTA	302
OY	422	ACTTTGTACCACTGGGATCTGCTCAGGCGCTTACGATTCGCTATGAGGCTGTGCTAAC	481
Db	303	ACCCCTTTCACTGGGATGTTCCGAGGCTTTTGAGAGATCGGTACGGGGGCAATGCTGAAC	362
OY	482	GTGGAAGAGGCTCAGCTGGACTTTGAGGGGATGTGGAGGTGTGTGTTGAAAGTTTGGG	541
Db	363	CAGGAGAGGTTCATTCCTGATCTTGTCGCTTACGGCGCTGTCTGTTTGAACGCGCTAAGG	422
OY	542	GACCGAGTCCAGAACTGGATCAACATCAACGAMCCCTGG-ATTACGGCGATCTATGATA	600
Db	423	CCAAAGTCTCGTCACTGATTTACATTCATGAAACATGTGTGTACTGCTGGCAGGTTA	482
OY	601	TGCCACCGGACAGCAACGCCCGGGCAGAGAGCAATTAAACAAGCACTCCACCGAGGGCA	660
Db	483	TGCGGCAAGTGTATATGCTCTCCGACGGTCACTGTTTGCAGCACTCAATGAAAGGGTGA	542
OY	661	CACGTGCCACTGAGCCGTGGCTGTGCTGGAAGGCGCAATCATAGCCATGCCCGGCGCT	720
Db	543	TTCTCTGACAGAACCGTTTCATGTGTGGCCACAACAAATGTGTGACCAACCGCGCATGTTTC	602
OY	721	GAGCGGTCTACAGCAGGGACTTTGGCGCCCTCGCAAAAGGGCCAGATCGGATCTCGTCAA	780
Db	603	CAGCTTTACCGAAGGAGGTGTTCAACCGGACGAAAGAAACATTTGGCATACCTTCCA	662
OY	781	CGGCGACTACTATGAGCCCTGGGACAGCAATGAGCCTCGGACACAGAGGCTGTGTAACG	840
Db	663	TGGCAACTGTGTGGAACTTTGGGATGAAAGAGATCGGGGGACAGAAAGCAGACGACG	722
OY	841	ACGGATGGAAATTTCACTTTGGCTGTTTCCAAATCCATCTTCTTGAAGAAAGTACTATCC	900
Db	723	GAGCGGGAATTCGAGATCGCTGTGTTCCCGGACCCCTTAATACAGACAGCGACTACCC	782
OY	901	AGAGAGCATGGAAGAGAGAGCTGGGGGAGAGGTTTCACCCCTCACTCCCGGCGACTTGGC	960
Db	783	AGCCTCAATGGGGGCTCAACTCGGGGAGCGTTTACCGGTTTCACTCCGAGGAGT---C	839
OY	961	CATCCTCATGCGCGAGAGACCGAATTTCTACGGCATGAAATTACTACATCCCAAGTTGCG	1020
Db	840	AAATCTGTATCTAGGAAGTTAGAAATTTTACCGTATGAACTCGTACAGACCTTTTGTGT	899
OY	1021	GCGCCACTTGAACGGTCCGTCCTCCGAGACGGAATTAATCTGGCGGCATTCATGAGCACCA	1080
Db	900	GCAGCACAAGGATAGCGCTCCAGATATCAATGACCAACAAAGCAATGTCAATGTTTACGA	959
OY	1081	GGAAGATTAAGACCGGCAACCCGTTTGGGAGAGAGGCGGCTCGCTGGCTGTGCTCTGT	1140
Db	960	TACAAACAGCAAAAGCGTCTCTCGAGGGGAAAGTCCGATACGCTTGGCTGCGGACGGC	1019
OY	1141	CCCGGACATGTTCCGGAAGCATCTGCGCCGGGTTATACGGCTGTACGGCAAGCCCATTA	1200
Db	1020	ACCCACTGGAATGGCGCAAGCTGTCTCAATTTGATTTGGAACCGATATCATGTGCTTATATA	1079
OY	1201	CATCACCGAGAACGGATGCCGTGCTCTGAGAGAGGAACAATGACGTGCGAGAGGCGCT	1260
Db	1080	TGTGACGGAAATATG--GACGAGACAGCAAAAGGAGAGACAGTCCACACAGAGGTACT	1136
OY	1261	CACGACCCCTTCCGCAATCCGTACTTTGATCTCGCACTTTGACTTCGATTTTCCAAAGGCAT	1320

2

```

Db      1137 CATGACACTTTCGGTAGCGATTCTTCGAAGGCTACGTTGTGGAATTGGCCCGTCTGT 1196
Qy      1321 TACCCAGGACGGCGTCTCGTCAAGGGGTACTTTGCGTGGGCGTTGCTCGATACTTGG 1379
Db      1197 CAAAGAGGATGGAGTGGATATCCGGTCTTACTTTGCCCTGGACCTTCACCGCAACTGGGG 1255

Search completed: March 29, 2004, 02:50:59
Job time : 6866.96 secs

```

2

LENGTH: 2016
TYPE: DNA
ORGANISM: Homo Sapien
US-10-152-372-253

Query Match
Best Local Similarity 54.5%; Pred. No. 6.9e-36;
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

```
QY 76 AGCTCTGCCCAACGACTTTGGAATGGGGCTTCGCAACGGCGCGCTACAGATCGAAGCGC 135
DB 237 AACCTCCCTCTTGGCTTCTCTGCGGGGTGGGAGGCTTCTACAGACGAGGGCGC 296
QY 136 CGTCAAGAAGGTGGCGCGCGCGCTTCATCTGGAGACAGTACCTGACCTGGAGCCATC 195
DB 297 CTGGGACCGAGACGGGAAGGGGCTTAGCATCTGGAGCTTTCACACACAGTGGGAAGG 356
QY 196 GCGACCAACGAG---CGCCAAACGGGATGTGGCTTGGCATCACTACACCGCTACGATGA 252
DB 357 GAAAGTCTTGGGATGAGACGGCAGATGTAGCCTGTGACGGCTACTACAAAGTCCAGGA 416
QY 253 GGACTTGTATCTTGAACAAAGTACGGCGCAAGGCTTCCGCTTCTCTTGTGTGTC 312
DB 417 GGACATCATCTCTGCTGAGGGAATGCACTGCAACCACTACCGATTTCTCTGTGTGCC 476
QY 313 GCGGATCATTTCCCTCGCGCGGAGGCTGGATCCGCTCAAGAGAGAGGGAATTGATTTTA 372
DB 477 CCGGCTCTGCTCCACAGGCAATCCAGCGGAGAGGTGAACAAGAGGAAATCGAATTTTA 536
QY 373 CAGCAAACTGATTAACGCTCTGTGAGGCGGGGATATCAAGCCTTGGTGACTTTGTACA 432
DB 537 CAGTGATTTTATGATGCTCTTCTGAGCAGCAATCATCTCCATCTGTGACCTTGTGAC 596
QY 433 CTGGGATCTGCTTGAAGCGCTTCAAGATCGCTATGAGAGCTGGCTCAACGTGGAAGGT 492
DB 597 CTGGGATCTGCTTGAAGCGCTTCAAGATCGCTATGAGAGCTGGCTCAACGTGGAAGGT 656
QY 493 CAGGTGAGCTTTGAGGCGGTATGCGAGGTTGCTTGAAGCTTTGGGAGACGAGTCCA 552
DB 657 CAA---CTACTTCAGAGACTAGCCCACTGTGCTTTTGAAGCCTTTGGGAGACGAGTAA 713
QY 553 GAACTGATCAACATCAAGCAAGWCCCTGTGATTCAGGCACTATGATATGCAACGGGAG 612
DB 714 GCACTGATCAAGTTCAGTGAATCTCGGCAATGCAAGAAAGGCTATGAGACGGGCA 773
QY 613 CAACGCCCCGGGAGAGAGAGATTAACAAGACTCCACGAGGGCAACACTGCACTGA 672
DB 774 CCATCGCGCGGGCTGGAAGCTCCG-----GGGACCGG 806
QY 673 GCGGTGGCTCGCTGGAAGGGCCAGATCATGAGCCGCGCGCGCTGAGCTTACAG 732
DB 807 CTTGTAACAAGGAGACACACATATTAAGGCCCAACGCAAAACCTGCACTTTTAATA 866
QY 733 CAGGAGCTTTCGCCCCCTCGCAAAAGGGCCAGATCGGCACTTGTCTCAACGGGCACTACTA 792
DB 867 CACCAAGTGGGCGCAGCAAGCAAGAGTCTGAGTGGAAATTTCACTGAATGTGAGCTGGG 926
QY 793 TGAGCCCTGGGACAGCATGAGCTTCGGGACAAAGAGGTGTGTGAGCAAGATGGAATT 852
DB 927 GGAACCTGTGACATTAATTAACCCCAAGGACCTTAGAGGCTGCGAGATACCTACAGTT 986
QY 853 TCACATTTGGCTGTTTGGCAATCCCATCT 881
DB 987 CTGTCTGGGCTGTGTTGCCAACCCCATTT 1015
```

RESULT 2
PCT-US04-02242-95
Sequence 95. Application PC/TUS0402242
GENERAL INFORMATION:
APPLICANT: Diversa Corporation
APPLICANT: Barton, Robertson, Dan; Elkins, James; Chang, Kristine
TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN

TITLE OF INVENTION: THEM
FILE REFERENCE: 56446-2018.40
CURRENT APPLICATION NUMBER: PCT/US04/02242
CURRENT FILING DATE: 2004-01-26
PRIOR APPLICATION NUMBER: 60/442,794
PRIOR FILING DATE: 2003-01-24
NUMBER OF SEQ ID NOS: 224
SOFTWARE: PatentIn version 3.1
SEQ ID NO 95
LENGTH: 1374
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-95

Query Match
Best Local Similarity 60.0%; Pred. No. 5.2e-32;
Matches 329; Conservative 0; Mismatches 207; Indels 12; Gaps 3;

```
QY 81 TGCCCAACGACTTTGATGGGGCTTGCACACGGCGCGCTACAGATGGAAGGCGCTCA 140
DB 38 TTCCCGCACTTTCACCTGAGGGGTGGCCACACAGCGCTTACAGATGGAAGGCGCGCG 97
QY 141 AAGAAGTGGCGCGCGCGCTTCATCTGGGACACGTACTGCCACTGAGACCATCGGCA 200
DB 98 CCAATGGCGCGCGCGCGCTTCATCTGGGATACCTTCAAGCCACACGGAAGGCAAGATCA 157
QY 201 CCAAGCGCGCAACGGGATGTGGCTTGGCATCACTACACCGCTAGATGAGAGCTTTG 260
DB 158 TCGACGGCACAATGGGAGGTGGCTTGCACACACTACACCGCTATGCGAGAGCTGG 217
QY 261 ATCTTGAACCAAGTACGGGCAAGGCTTACGCTTCTCTTGTGTGTCGCTGGATCA 320
DB 218 AGCTGATCGCAGCTTGGGCTGAGAACCTTACAGCTTTTCAATGTCCTGGCGCGCTCC 277
QY 321 TTCCCTCGCGCGGAGCTGATCCGTCAACGAGAGGGAATTGATTTTACAGCAAC 380
DB 278 AGCCCAAGGGGTCC-----GGCGCTGGAACGAAGAGGCTTGAATTTCTATGCGCGC 331
QY 381 TGATTTGAGCGCTTGTGAGGCGGGGATTAACGCTTGGTGTGATCTTGAACCTGGATC 440
DB 332 TGTCTGAGCGCTTGGCGCGCAAGGACTCAAGCGACCTGACCTTATCCACTGGGAC 391
QY 441 TGCTCAGGCGCTTACAGATGCTATGAGAGGCTGCTCAACGTTGGAAGAGGTCCAGCTGG 500
DB 392 TGCCGCAAGCTTTCAGAGA---CGAGGGCGCTGTGCTCAATCGGCAACCTGTACC--- 445
QY 501 ACTTTGAGCGGTATGCGAGTGTGCTTTGAAGCTTTTGGGAGCCAGTCCAGAACTGGA 560
DB 446 ACTTGCAGCGATGATGCGCGCGAGGTTGGCGCGCTTTCGCGCAAGGTGCGCAGATCG 505
QY 561 TCACATCAACAGACCGCTGATTCAGGCACTATGATGATGACCGGCGAGCAACGCC 620
DB 506 CCAGCAACATGAGCCGTGTGACCTGCGTGTGGGCAAGGACCGGCAAGTTCGCGC 565
QY 621 CCGGCAAG 628
DB 566 CCGGCAAG 573
```

RESULT 3
US-10-767-701-7585
Sequence 7585. Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29

NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 7585
; LENGTH: 805
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: clone ID: SORBI-28MAY03-CLUS93553_1
; US-10-767-701-7585

Query Match 8.0%; Score 159.8; DB 6; Length 805;
Best Local Similarity 54.3%; Pred. No. 2,4e-31;
Matches 345; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

QY 81 TGGCCACGACTTGAATGGGCTTCGCAACGGCCGCTACCATGGAAGCGCCCTCA 140
DB 174 TCCCAAGAGCTTGTGTTGGGACAGGCTCCGGGCTATCATAGAGGTTGCTTACA 233
QY 141 AAGAAGTGGCCGCGCCGCTCCATCTGGGACAGTACTGCACTGGAAGCCATCGCGCA 200
DB 234 ACGAAGAGGCAAGGGGCTTACCATATGGGACAAAGTTCACTACATCCAGTAAAGATCT 293
QY 201 CCAAGCGGCGCAAGGGGATGGCTTGGATCATCTACACCGCTAGAGTGAAGACTTTG 260
DB 294 TGAACAACGATACCGGCGAGCTAGCAGATGACATGATCAATCAAGAGAGATGTC 353
QY 261 ATCTTGAACCAAGTACGGGCAAGGCTTACCGCTTCTCTTGTCTGGTTCGGGATCA 320
DB 354 AACTCTCAAGGACATGAACTTGGAGCCCTTCGGCTTCTCATTTGCTGGAACAGATCC 413
QY 321 TTCCTCTGGCGGAGGCTGATCCGCTCAACGAGGGAATTGATTTTACAGCAAC 380
DB 414 TGGCAACGGTTCCTGAGTGGAGGAATCAACAAAGAGGCTGGCTCTTATTAACAAC 473
QY 381 TGAATGACGGCTTGAAGCGGGGTATCAAGCTTGGGTGATTTTACCATCTGGATC 440
DB 474 TCAATCAACGAGTATAGCCAAAGGCTGAAGCCATTGTGATCCATCTTCACTGGGACA 533
QY 441 TGGCTGAGGCTTCAAGATCGCTAGAGGCTGGCTCAACGTGGAAGAGTCCAGCTGG 500
DB 534 CGCCCTCGGCTTGGAGGACAGATGAGATTTCTCAGCGAAGATCATCA---GG 590
QY 501 ACTTTGAGCGTATGCGAGGTGTGCTTTGAACGTTTGGGACCGAGTCCAGAACTGA 560
DB 591 ACTAGGAGACTTGGCGAGGTTTGTCTTCAAGAGTTTGGCGACCGGTCAAGGCTGGA 650
QY 561 TCACCATCAACGAGCCCTGATTCAGGCTATGATATGCAACGGGACAGCAAGCCCC 620
DB 651 CCAAGTTCAACGAGCCCTGACGTAATGACCAAGGCTAGCGCTGGCAAGTCCGCGC 710
QY 621 CGGGCAGGACAGCATTAACAAGACTCCACGAGGGCAACATGCACTGAGCGGTGGC 680
DB 711 CGGGCGGCTGTGTGTCTGATGCTCAACAAAGATGCTTCCCGGCGACTCGGCGGAGC 770
QY 681 TCGCTGAAAAGGCCAGATCATGAGCCATGCGCGC 715
DB 771 CTTACACCGTGAAGCAACATCATCTCTGCCCCAC 805

RESULT 4
US-10-767-471-414
; Sequence 414, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 6274
; TYPE: DNA

ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(6274)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-;
; US-10-767-471-414

Query Match 7.6%; Score 151.6; DB 6; Length 6274;
Best Local Similarity 52.4%; Pred. No. 7,4e-29;
Matches 446; Conservative 1; Mismatches 371; Indels 33; Gaps 4;

QY 81 TGGCCACGACTTGAATGGGCTTCGCAACGGCCGCTTACCATGGAAGCGCCCTCA 140
DB 4141 TTTCTGAGGCTTCACTGGAAGGAGGAGCTTCTCTCATATCAGATTGAAGGCTTGA 4200
QY 141 AAGAAGTGGCCGCGCCGCTCCATCTGGGACAGTACTGCACTGGAAGCCATCGCGCA 200
DB 4201 GAGCAGATGGCAAGAGACTGACATTTGGGACAGTTTCTTCAACACCACTGAGGTTG 4260
QY 201 CCAAGCGGCGCAAGGGGATGGCTTGGATCATCTACACCGCTAGAGTGAAGACTTTG 260
DB 4261 AGAAGATGCTATGGAAGAGTGGCTCTGGAAGTTATCAAGATTTGCTGAGGATCTGG 4320
QY 261 ATCTTGAACCAAGTACGGGCAAGGCTTACCGCTTCTCTTGTCTGGTTCGGGATCA 320
DB 4321 TCACCTGCAAGACTGGGGTGTCCACTACCGTTTTCATCTCTGCTGCTGCTATCC 4380
QY 321 TTCCTCTGGCGGAGGCTGATCCGCTCAACGAGGAGGAATTGATTTTACAGCAAC 380
DB 4381 TCCCTGATGACACCAAGGTA--CATCAATAAAGCGGCTGAACTACTAGTGGGCT 4437
QY 381 TGAATGACGGCTTGTGAGCGGGGTATCAAGCTTGGGTGATTTTGAACATCTGGATC 440
DB 4438 TCAATGATACATGCTGCGCGGACGATCCAGGCCAGGATTAATTAACATGGGACC 4497
QY 441 TGGCTGAGGCTTCAAGATCGCTATGAGGCTGCTCAACGTGGAAGAGTCCAGCTGG 500
DB 4498 TACCAAGAGCGCTCAAGAT---GTAGAGGCTGGGAGAAATGAGACATCTGTCAGCGG- 4553
QY 501 ACTTTGAGCGTATGAGAGTTGTGCTTTGAACGTTTGGGACCGAGTCCAGAACTGA 560
DB 4554 -TTTAAGGAGTATGAGATGTGCTTTCAGAGGCTGGGAGCAAGTGAATTTTGA 4611
QY 561 TCACCATCAACGAGCCCTGATTCAGGCTATGATATGCAACGGGACAGCAAGCCCC 620
DB 4612 TCACGTTGAATGAGCCCTTTGTATGCTTATGCAAGGCTATGCTACGGAACAGCAGCTC 4671
QY 621 CGGGCAGGACAGCATTAACAAGACTCCACGAGGGCAACATGCACTGAGCGGTGGC 680
DB 4672 CAGGAGTCTCCA-----ATAGGCTTGGCACTGCCCCCTACA 4707
QY 681 TCGCTGAAAAGGCCAGATCATGAGCCATGCGCGCCGCTGTACTAGAGGAGAT 740
DB 4708 TTTGTGGCCACAAATCAATTAAGGCTCATGCTGAGGCTTGGCATCTGTACAGATGTGT 4767
QY 741 TTTGCGCCCTGCAAAAGGCGCAGATCGGATCTCGCTCAACGGCGACTATAGAGCCCT 800
DB 4768 ACCGCGCAGTCAAGGTGGCTGTATTTCTATCACTACAGAGTATGCGGCTGAACCCA 4827
QY 801 GGGACAGCAATGAGCTCGGAGCAAGAGGCTGTGAGCGAGATGGAATTTTCAATTG 860
DB 4828 GAGATCCCTTCAACGAGAGGATGTGAGGACCGAGAGATATGTTCATGATTCATGGAG 4887
QY 861 GCTGTTTGGCAATCCCATCTTTTGAAGAAGACTATCAAGAGACTGAAGAAGACAGC 920
DB 4888 GCTGTTTGGCAATCCCATCTTTTGAAGAAGACTATCAAGAGACTGAAGAAGACAGC 920
QY 921 TGGCGAGAGG 931
DB 4948 TCCGTGACAGG 4958

RESULT 5

```

RESULT 6
PCT-US04-02242-93
; Sequence 93 Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Divetisa Corporation
; APPLICANT: Batton, Nelson, Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING ANTI
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 56446-20118_40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; CURRENT FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 93
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-93

```

Query	March	7.0%	Score 139;	DB 1;	Length 1389;
Best Local Similarity	57.1%	Pred. No. 6,4e-26;			
Matches	316;	Conservative 0;	Mismatches 225;	Indels 12;	Gaps 3;
QY	72	CGCTAGCTCTGCCCAACGACTTTGAATGGAGGCTTCGCAACGCGCCGCTACCGATCGAAG	131		
Db	26	CCCTGTCCTTCCAGAGCAGTTGCTGTCGGGGTGTCTGCCGGGCTCTTACCAAGTTCAGG	85		
QY	132	GGCGCGTCAAGAAGAGGTGCGCCGCGCCGTCATCTGGGACACGTACTGCCACTCGAGC	191		
Db	86	GGCCGCTTCCACGAGACGGGGAAGGGCCCTTCGCTTGGGACATGTTCTGCAAGAAGCCG	145		
QY	192	CATCGGCGACCAACGGCGCCAAACGGCGATGTGTGCTTTCGATCACTACCAACCGCTACGATG	251		
Db	146	GAGCGGCTTTCACAGGAGGACGACGCGGGGCGGTGGCTTGGACCACTATCACCGTACCGAG	205		
QY	252	AGGACTTTGATCTCTTGAACAAAGTACGCGGCAAAAGGCTTACCGCTTCCCTGTGCGGT	311		
Db	206	AGGACGTGGCGTTGATGCGACAGGTGGGCTGCAAGCGCTAACCGGCTAAGCTGTGCTGGC	265		
QY	312	CGCGATCATTTCCCTCGGCGGCGAGGCTGATATCCCGTCAACGAGAGGGAATTAGTTTT	371		
Db	266	CCCGAGTGTCTCCGAGGGGCGTCCGGCAG-----CCCAACGAAAGAGGCTCGACTTCT	319		
QY	372	ACAGGAACCTGATTAACGCGCTGTTTGAAGCGGGGATTAACGCGCTTGGGTACTTTTGACC	431		
Db	320	ACTGCGGTTGTGTAGACGCGCTGCTTCAGGACAGGGAATTAAGCCCTGGGTAAAGCTTTTTC	379		
QY	432	ACTGGGATCTGCTTCAGGCGGCTTACAGATGCTATGGAAGGCTGGCTCAACGTGGAAGAG	491		
Db	380	ATTGGGACTACCCCTTGCTCT---CTATACCGGGGGGCGCTTGCTCAACCGGGAATAGG	436		
QY	492	TCCAGCTGACCTTTGACACGGTATGCCAAGTTGTGCTTTTAAACGTTTGGGGAACGAGTCC	551		
Db	437	CGGATTTGG---TTTCCGAGTACGCGGGCTTAATCCCGATGCGCTCTCCGACCGGGTGC	493		
QY	552	AGAACTGATACCATCAACGAAAGGCTGTTCAGGCTCACTGATGATATGGCACCCGGA	611		
Db	494	AGCAATTTCTTCACTCAAGAACGAGCCTCCAGGTCTATATCGGCTTGGACACCTCGAGGGTA	553		
QY	612	GCAACGCCCCGGG	624		
Db	554	AGCATGTCTCAGG	566		

APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 FILE REFERENCE: 38-21(5353)B
 CURRENT APPLICATION NUMBER: US/10/767,701
 CURRENT FILING DATE: 2004-01-29
 NUMBER OF SEQ ID NOS: 63128
 SEQ ID NO 14696
 LENGTH: 2042
 TYPE: DNA
 ORGANISM: Sorghum bicolor
 FEATURE:
 OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS4596_1
 US-10-767-701-14696

Query Match 6.4%; Score 131; DB 6; Length 2042;
 Best Local Similarity 51.5%; Pred. No. 8.5e-24;
 Matches 433; Conservative 1; Mismatches 386; Indels 21; Gaps 5;

81 TGGCCAAACGACTTGAATGGGGCTTGGCAACGGCCGCTTCCAGATGAGGCGCCCTCA 140
 165 TCCCGCGGGGCTTCTTCTGGCGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 224
 141 AAGAAGTGGCGCGCGCGCGCTTCCATCTGGGACAGTACTCCACCTGAGCCATCGCGCA 200
 225 CAGAGATGGAAGGAAGGAGCTTGGGACACATTCACACATTAAGATATTTCT--A 281
 201 CCAACGGCGCAACGGGATGCGCTTGGCATCACTACCAACCGCTACGATGAGGACTTTG 260
 282 TTGACGGTGGCAAGGATGATTAAGTCCGATCAGTATCAATGAATCAAGGAAGATGTA 341
 261 ATCTTGAACCAATGACGGGCAAGGCTTACCGCTTCTCTCTCTCTCTCTCTCTCTCTCT 320
 342 AGCTTTTGAATGAGTGGGCTTGGATGCTTACAGAGATGATGATGATGATGATGATGAT 401
 321 TTCCCTCGCGGCGGATGATCCGCTCAACGAGAGGGAATTTGATTTTACAGCAAC 380
 402 TTCTCTATGCGC-----CGAGAGCTGTCAATCCGAAGGGGCTGGATCTACACATC 455
 381 TGATTAAGCGCTTGTGAGCGGGGATACAGCCCTTGGGATCTTTGATCACTGGGATC 440
 456 TGATTAAGTAATCTTATGCTTACGAGTAATCACTCAATGATATCATTTGATTTGAT 515
 441 TGGCTCAGGCGCTTCAAGATCCGCTATGAGAGGCTGGCTCAACGATGAGAGTCCAGTGG 500
 516 TTCTCAGGCTCTTCAAGATGAGTACAAAGGGGCTGTTAGTCTTGA--TTGATAGAG 572
 501 ACTTTAGCGGATGAGAGGCTTGTGATCGTTTGAACGTTTGGGAGCCGATCCAGAACTGA 560
 573 ACTACACAGATACGAGATGTGTCTTCAAGAACTTTGGCGAGAGTGAAGTACTGGA 632
 561 TCACCATCAACGAGCCCTGATTCAGGCACTATGATATGCAACCGGAGCAACCGCC 620
 633 GCACTGCAATGAACTTAATCAAGATGAGTGGCGGATTTGACCAAGAACTCTACAC 692
 621 C-----GGGCAAGAGATTAACAAGCACTCAACGAGGGAACAACCTCCACTAGC 674
 693 CACAGGAGATCTAGCCCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGG 752
 675 CGTGGCTCGTGGAAAGGCGCAATCATGAGCAATGCGCGCGCTGAGGCGCTGATCAAGCA 734
 753 CGTACATGAGACACCACTTGTACTTGGCGATGATCTGCAAGTCTCCCTATATAGAG 812
 735 GGAAGCTTGGCGCTTGGCAAAAGGCGCAATCGGCACTTGGCTCAACGCGGATCAATG 794
 813 AGAAGTACAGGCTGAGCAAGAGAGCAATGCAAGCGCTTGGCTTGGGCTGATG 872
 795 AGCCCTGGAGACAGATGAGCTTGGGCAAGAGGCTTGGGCAAGCAAGATGAATTTTC 854
 873 AGCTT--GCAACGAGAGCTTCCGATATATTTGAGAGGCTGCAAGGATGAATATTTTC 929
 855 ACATTGGCTGTTTGGCAATCCCATCTTTTGAAGAGCACTATCCAGAGCATGAGA 914

Db 930 ACCTGAGTGGTTCATGATCTCTATGATGAGATACCTCCAGTATGAGCAAGA 989
 Qy 915 A 915
 Db 990 A 990

RESULT 8
 US-10-093-037A-2
 Sequence 2, Application US/10093037A

GENERAL INFORMATION:
 APPLICANT: Jay M. Short
 APPLICANT: Yihua, Edward
 APPLICANT: Swanson, Ronald V.
 APPLICANT: Mathur, Eric J.

TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
 FILE REFERENCE: 564462001402
 CURRENT APPLICATION NUMBER: US/10/093,037A
 CURRENT FILING DATE: 2002-03-06
 PRIOR APPLICATION NUMBER: US 09/910,579
 PRIOR FILING DATE: 2001-07-20
 PRIOR APPLICATION NUMBER: US 09/134,078
 PRIOR FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 08/949,026
 PRIOR FILING DATE: 1997-10-10
 PRIOR APPLICATION NUMBER: US 60/056,916
 PRIOR FILING DATE: 1996-12-06
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2
 LENGTH: 1317
 TYPE: DNA
 ORGANISM: Thermotoga sp.
 US-10-093-037A-2

Query Match 6.4%; Score 130.8; DB 6; Length 1317;
 Best Local Similarity 56.4%; Pred. No. 7.9e-24;
 Matches 309; Conservative 1; Mismatches 223; Indels 15; Gaps 3;

77 GCTCTGCCCAACGACTTTGATGAGGCTTGCACAGCGCGCTTACAGATGAGGCGCC 136
 16 GATTTTCCAAAGATTTTATCTTGGACGGCTTACCGACATACAGATGAGGAGTGA 75
 137 GTCAGAGAGTGGCGCGCGCGCTTCACTTGGGACAGTACTGCCACTGGAGCCATCG 196
 76 GCAAGAGAGATGCGAGAGGCGCATCAATTTGGGATGCTTTTACACACGCGCTGGCAA 135
 197 CGCAGCAGCGCGCGCGAGTGGCTTGGATCACTACACCGCTGAGTGAAGAGAC 256
 136 ACCCTGAGCGGTACACAGAGAGCTTGGCTGATCACTTATACAGATCAAGAGAGAT 195
 257 TTTGATCTTGGACCAAGTACGCGCGCAAGGCTTACCGCTTCTCTGTGTGTCGCG 316
 196 ATCCAGGTATGAAGAAATAGGTTTACGCTTACAGGTTCTCTATCTCTGCGCCAGA 255
 317 ATCATTCCTTGGCGCGCGAGCTGATCCGCTCAAGAGAGGAAATGAGTTTACAC 376
 256 ATTATGCGAGATG-----GAAGAACATCAACCAAAAGGAGTGTGATTTCTACAC 306
 377 AAACGATGAGCGCGCTTGGAGCGGGGATGACGCGCTTGGGATCTTGTACACTGG 436
 307 AGACTGTTATAGACTTTTGAAGAAATGATATATACATTCGTAACACTCTTACTGG 366
 437 GATCTGCTCAGCGCTTCAAGATCGCTATGAGAGGCTGCTCAACGTAAGAGGTCAG 496
 367 GACTTACCTTACGACATTTATGAAA--AAGGTGATGAGTTAAACC--AGATATAGCG 420
 497 CTGACCTTGAAGGAGTATGAGAGTGTGCTTTGAACGTTTGGGAGCGAGTCCAGAAC 556
 421 CTCTATTTAGAGCAATACGCAAGCTTATGTTTCAAGAACTCGGTATGCTGTGAACAT 480

APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 12773
LENGTH: 716
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS25405_1
US-10-767-701-12773

Query Match 5.8%; Score 115.4; DB 6; Length 716;

Best Local Similarity 56.6%; Pred. No. 5.3e-20; Mismatches 176; Indels 6; Gaps 1;

Matches 237; Conservative 0; Mismatches 176; Indels 6; Gaps 1;

QY 81 TGCCCAAGACTTGAATGGGGCTTCGCAAGCGCCGCTCAAGATCGAAGCGCCGCA 140
DB |||||
DB 281 TCCCAAGGGGTGTGTGGGAGCGGAGCTGGGCTACAGGTGAGGGCCGCGT 340
QY 141 AAGAGGTGGCGCGCCGCTTCATCTGGACACGTACTGCCCTGAGCCATCGCGCA 200
DB |||||
DB 341 CCACCAAGCGCGGAGACCTTCATCTGGATTCATTCGCGCAGTCCAGAAATATG 400
QY 201 CCAAGCGCGCAAGCGGATGCGCTTGCATCTACCAAGCTAGATAGACTTTG 260
DB |||||
DB 401 CGGGAATCAAAATGAGACCTTGATGATCAATCAATCAATCAATCAATCAATCA 460
QY 261 ATCTTGACCAAGTACAGCGCCCAAGGCTTCCCTTGTGCTGTGGGATCA 320
DB |||||
DB 461 ATCTGATGAAGATTTGATTTGATGCTACCGTTTCAATCTCGTGTCCAGACT 520
QY 321 TTCCCTTGGCGGAGCTGATCCCGTCAAGAGAGGAAATGATTTTACAGCAAC 380
DB |||||
DB 521 TCCGAGATGGCG-----AGGGAAGATCAATCCAGAGGTGATATTAATAATC 574
QY 381 TGATGACCGCTTGTGAGGGGTATCAGCGCTGGGTATCTTGATGACATCGGATC 440
DB |||||
DB 575 TGAATTAATATCTCTTCAAGAGCATGACTCTTACATCAACCTTTACCATATGATC 634
QY 441 TGCTCAGCGCTTACAGATGCTATGAGGCTGCTCAAGTGAAGAGTCCAGCTG 499
DB |||||
DB 635 TTCTCTTGGCTTGAAGAAATATGAGGCTTAAAGGCTTAAGATGCGAGCTG 693

RESULT 14

US-10-767-701-26695
Sequence 26695, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 26695
LENGTH: 637
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: 30977777
US-10-767-701-26695

Query Match 5.8%; Score 114.8; DB 6; Length 637;
Best Local Similarity 58.4%; Pred. No. 7.2e-20;
Matches 222; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 77 GCTTGGCCAAAGACTTTGAATGGGGCTTGGCAAGCGCCGCTTACAGATCGAAGCGCC 136
DB |||||
DB 256 GCGTCCCAAGGGGTTCATCTTCGGGACGGCAGCGTCCCGCTTCAGGTGAGGCGCG 315
QY 137 GTCAAGAAGGTGGCGCGCGCGCTGCATCTGGGACCGTATGCGCCAGCTGAGGCTCG 196
DB |||||
DB 316 GCACGTCGCGCGCGCGCGCGCGCATCTGGGATCTCTTCGTCACACCCCGGAAG 375
QY 197 CGACCAAGCGCGCGCGCGCGCGATGCTTGGATCTACCAAGCTGATGAGAC 256
DB |||||
DB 376 ATGCTGAAGACGGGAACGAGATGTACAGATGAATATATCTGCTCAAGAGAT 435
QY 257 TTGATCTCTTGAACCAAGTACGGCGCAAGGCTTACCGCTTCTCTTGTGTGCGCG 316
DB |||||
DB 436 GTTCATCTCATGAAGAACCTTAATTTGATGACATACCGGTTTCAATCTCTGCTCAGG 495
QY 317 ATCATTCCTCGCGCGCGCGCGCGTGCATCCGCTCAAGAGGAGATTTAGAGC 376
DB |||||
DB 496 ATCTTCCAGATGCGGAAGG-----AAAGTATGAAGAGGATGACGATTTACAC 549
QY 377 AAAGTATGACCGCGCTTGGAGCGGGGTATCAGCGCTTGGGTGATCTTGTACCACTGG 436
DB |||||
DB 550 AATCTTATGAATCAATGATTAAGCAAGTCTTACTCTTACGCCAATTAACCATAT 609
QY 437 GATCTGCTGAGGCGCTTCA 456
DB |||||
DB 610 GATCTTCACTTGGGCTTCA 629

RESULT 15

US-10-767-795-4474
Sequence 4474, Application US/10767795
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,795
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 117596
SEQ ID NO 4474
LENGTH: 558
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C30142_1
US-10-767-795-4474

Query Match 5.5%; Score 109; DB 6; Length 558;
Best Local Similarity 58.7%; Pred. No. 2.1e-18;
Matches 225; Conservative 0; Mismatches 155; Indels 3; Gaps 2;

QY 77 GCTTGGCCAAAGACTTTGAATGGGGCTTGGCAAGCGCCGCTTACAGATCGAAGCGCC 136
DB |||||
DB 107 GATTTCCCTCCCACTTGTCTTGGGTTCATCTTCTTCACTTCTTCAAGATGAGAGGT 166
QY 137 GTCAAGAAGGTGGCGCGCGCGCGCTGCATCTGGGACCGTATGCGCCAGCTGAGGCTCG 196
DB |||||
DB 167 GTGAACGAAGGTGCAAGAGTAAAGATATGGGATCTTCTTCTCAATTTGAAGAAA 226
QY 197 CGACCAAGCGCGCGCGCGCGCGATGCTTGGATCTACCAAGCTGATGAGAC 256
DB |||||
DB 227 ATGCTGATGAAGCAATGCTGATGTTGCACTGATCTTACCAAGATGAGAGAT 286
QY 257 TTGATCTTGAACCAAGTACGGCGCAAGGCTTACCGCTTCTCTTGTGTGCGCG 316
DB |||||
DB 287 ATAGAGCTTATATCGATGATGAGGTCAAGCTTACAGATTTTCAATATCATGCGCTCGT 346
QY 317 ATCATTCCTCGCGCGCGCGCGCGTGCATCCGCTCAAGAGGAGATTTAGAGC 376
DB |||||
DB 347 ATTTTCCAGATG--GTTTGGAAACCAAGATTAAGAGGAGATTTTACAC 404

Qy	377	AAACTGATTGACGCCCTGTGAGGGGGTATCACGCCCTTGGGTGACTTTGTACCACTGG	436
Db	405	AATGTGATCGATGCCCC-TTTGAAAGGTAATTGAGCCCTTTGTGACATTATACCAATTGG	463
Qy	437	GATCTGCCCTCAGGGGCTTTCACGA	459
Db	464	GATTCTTCCTTGCATCTCCATGA	486

Search completed: March 29, 2004, 02:56:06
Job time : 176.02 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 20:57:07 ; Search time 3903.5 Seconds
(without alignments)
11130.850 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455
Sequence: 1 atgcccagctcgtcgtctc.....gggttaagtggcgcatcaa 1455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est4:*
14: gb_est4:*
15: em_estfun:*
16: em_estfun:*
17: em_estfun:*
18: em_estfun:*
19: em_estfun:*
20: em_estfun:*
21: em_estfun:*
22: em_estfun:*
23: em_estfun:*
24: em_estfun:*
25: em_estfun:*
26: em_estfun:*
27: em_estfun:*
28: gb_est1:*
29: gb_est2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	702.2	48.3	773	14	CB904679 trico39xe
2	701	48.2	703	14	CF875727 trico39xe
3	461.2	31.7	584	14	CF884351 trico39xe
4	387	26.6	1994	9	AA415086 Mg0020 RC

5	201	13.8	2031	11	AY109400
6	195.2	13.4	786	14	CB677158
7	193	13.3	788	14	CB673495
8	190.4	13.1	806	14	CB642881
9	179.2	12.3	889	14	CD438244
10	175.6	12.1	480	14	CB688082
11	175	12.0	1182	11	AY106991
12	171.8	11.8	662	14	CF880605
13	171.8	11.8	752	14	CB907493
14	169.8	11.7	650	10	BE777259
15	162.6	11.2	613	9	AJ436490
16	158.6	10.9	860	14	CF666831
17	157.8	10.8	610	13	BQ110253
18	157	10.8	825	14	CK200642
19	156.2	10.7	1126	14	CK209455
20	155.8	10.7	708	13	CA145286
21	155.6	10.7	673	14	CD482402
22	155.6	10.7	674	14	CD458246
23	154	10.6	786	14	CB634193
24	153.4	10.5	792	14	CB633739
25	153.4	10.5	808	14	CB633785
26	153.2	10.5	739	13	CA139881
27	153	10.5	1126	14	CK161716
28	152.4	10.5	788	14	CF554428
29	151.8	10.4	728	14	CF473575
30	151.6	10.4	723	12	BM605139
31	151.6	10.4	766	14	CB643857
32	151.6	10.4	821	14	CB634605
33	151.2	10.4	564	14	CB641046
34	151.2	10.4	636	9	AV913799
35	151	10.4	676	12	BM621029
36	151	10.4	805	14	CB677500
37	150.8	10.4	646	12	BJ235347
38	150.8	10.4	839	14	CK201813
39	150.6	10.4	829	14	CB663657
40	150	10.3	592	14	CA193361
41	150	10.3	760	14	CB643189
42	150	10.3	794	14	CB654045
43	150	10.3	806	14	CB674932
44	150	10.3	826	14	CB648777
45	150	10.3	856	14	CB668311

ALIGNMENTS

RESULT 1
LOCUS CB904679
DEFINITION trico39xe13 T. reesei mycelial culture, Version 3 apr11 Hypocrea
ACCESSION CB904679.1 GI:30119337
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
FOREMAN, P.K., BROWN, D.E., DANKMEYER, L., DEAN, R., DIENER, S.,
DUNN-COLEMAN, N.S., GOEDEGEBUUR, F., HOUFEK, T.D., ENGLAND, G.J.,
KELLEY, A.S., MEERMAN, H.J., MITCHELL, T., MITCHINSON, C.,
OLIVARES, H.A., TEUNISSEN, P.J., YAO, J. and WARD, M.,
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus *Trichoderma reesei*
J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE
PUBMED
COMMENT
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817

Email: Pforeman@genecor.com
Seq primer: LT-F1 primer.
Location/Qualifiers
source

1..773
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric039x13"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/note="Vector: PREP3; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

```

Query Match      48.3%; Score 702.2; DB 14; Length 773;
Best Local Similarity 92.8%; Pred. No. 5.1e-129;
Matches 707; Conservative 1; Mismatches 54; Indels 0; Gaps 0;

279  GATCCCGCTCAACGAGGAGGAATTGATTTACAGCAACTGATGACGCCCTGTTGAG 338
12  GNNNGCTGCTGTAATTAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 71
339  GCGGGGATATCAAGCCTTGGTGGTACTTTGTACCTGGGATCTGCTCAGCGCTTCA 398
72  GCGGGGATATCAAGCCTTGGTGGTACTTTGTACCTGGGATCTGCTCAGCGCTTCA 131
399  TCGCTATGAGAGCTGGCTCAACGTGGAAGAGTCCAGTGTGACTTTGAGCGGTATG 458
132  TCGCTATGAGAGCTGGCTCAACGTGGAAGAGTCCAGTGTGACTTTGAGCGGTATG 191
459  GTTGTGCTTTGAAGCTTTGGGGACCGAGTCCAGAACTGATCAACATCAAGMCCCTG 518
192  GTTGTGCTTTGAAGCTTTGGGGACCGAGTCCAGAACTGATCAACATCAAGMCCCTG 251
519  GATTGAGGCTATGATGATATGCCACCGGACGACGCCCGGGGACAGACAGCATTTAA 578
252  GATTGAGGCTATGATGATATGCCACCGGACGACGCCCGGGGACAGACAGCATTTAA 311
579  CAAGCAGCTCCAGCGAGGACACCTGACCGGCTGCTGCTGGAAGGCCCAAGT 638
312  CAAGCAGCTCCAGCGAGGACACCTGACCGGCTGCTGCTGGAAGGCCCAAGT 371
639  CATGAGCAGTCCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 698
372  CATGAGCAGTCCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 431
699  CCAAGTGGGATCTTGGCTCAACGGGAGTACTATGAGCCTGGGACAGCAATGAGCCTG 758
432  CCAAGTGGGATCTTGGCTCAACGGGAGTACTATGAGCCTGGGACAGCAATGAGCCTG 491
759  GGAAGAAGAGGCTGCTGAGGAGCGGATGGAATTCACATTTGGCTGCTGCTGCTGCT 818
492  GGAAGAAGAGGCTGCTGAGGAGCGGATGGAATTCACATTTGGCTGCTGCTGCTGCT 551
819  CTTCTTGAAGAAGACTATCCAGAGACATGAGAGCAGTGGGCGAGAGGCTTCCAGC 878
552  CTTCTTGAAGAAGACTATCCAGAGACATGAGAGCAGTGGGCGAGAGGCTTCCAGC 611
879  CCTTACTCCCGCGAGCTTGGCATCTTCATGCTGCGGAGAGACCGACTTTCAGGATGAA 938
612  CCTTACTCCCGCGAGCTTGGCATCTTCATGCTGCGGAGAGACCGACTTTCAGGATGAA 671
939  TTACTACATCCAGTTCGCGGCGCACTAGACGATCCGCTGCCGAGAGCGGACTTACT 998
672  TTACTACATCCAGTTCGCGGCGCACTAGACGATCCGCTGCCGAGAGCGGACTTACT 731
999  CGGCGCATCATGAGACCAAGAGAAATAAGACGCGAGCCC 1040
732  CGGCGCATCATGAGACCAAGAGAAATAAGACGCGAGCCC 773

```

RESULT 2
CF875727
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CF875727 703 bp mRNA linear EST 31-OCT-2003
tric039x13.b1.T.reesei mycelial culture, Version 6 October 2003
Hypocrea jecorina cDNA clone tric039x13, mRNA sequence.
CF875727.1 GI:38130409
EST.

REFERENCE

Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
1 (bases 1 to 703)
Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,
Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
Dean, R.A.

AUTHORS

TITLE

Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset

JOURNAL

Unpublished (2003)
Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu

COMMENT

Seq primer: LT-F1 primer.
Location/Qualifiers

FEATURES

source

1..703
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric039x13"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October 2003"
/note="Vector: PREP3; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

```

Query Match      48.2%; Score 701; DB 14; Length 703;
Best Local Similarity 99.7%; Pred. No. 8.5e-129;
Matches 701; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

338  GGGGGGATATCAAGCCTTGGTGGTACTTTGTACCACTGGGATCTGCTCAGGCGCTTCA 397
1  GCGGGGATATCAAGCCTTGGTGGTACTTTGTACCACTGGGATCTGCTCAGGCGCTTCA 60
398  ATGCTATGAGAGGCTGGCTCAACGTGGAAGAGTCCAGCTGACTTGTGAGCGGTATGCGA 457
61  ATGCTATGAGAGGCTGGCTCAACGTGGAAGAGTCCAGCTGACTTGTGAGCGGTATGCGA 120
458  GGTGTGCTTTGAAGCTTTTGGGACCGAGTCCAGAACTGATGACATCAACGAGMCCCT 517
121  GGTGTGCTTTGAAGCTTTTGGGACCGAGTCCAGAACTGATGACATCAACGAGMCCCT 180
518  GGAATGAGGCACTTATGATATGCAACCGGAGCAACCCCGGGGACAGAGAGATTA 577
181  GGAATGAGGCACTTATGATATGCAACCGGAGCAACCCCGGGGACAGAGAGATTA 240
578  ACAAGACTCCACCGAGGCAACACTGCACTGAGCCGTGCTCGTGGAAAGGCCCA 637
241  ACAAGACTCCACCGAGGCAACACTGCACTGAGCCGTGCTCGTGGAAAGGCCCA 300
638  TCATGAGCAATGCGCGCGCTGCGCTTCAAGAGGACTTTGCGCCCTGCAAAAG 697
301  TCATGAGCAATGCGCGCGCTGCGCTTCAAGAGGACTTTGCGCCCTGCAAAAG 360
698  GCGAGATGGCATCTGCTCAACGAGGAGTACTATGAGCCCTGGGACAGCAATGAGCCTTC 757
361  GCCAGATGGCATCTGCTCAACGAGGAGTACTATGAGCCCTGGGACAGCAATGAGCCTTC 420

```


QY 758 GGGACAAGAGGCTGCTGAGCGACGAGATGGAATTTTCACATTGCTGTTTGGCAATCCCA 817
| | | | |
Db 421 GGGACAAGAGGCTGCTGAGCGACGAGATGGAATTTTCACATTGCTGTTTGGCAATCCCA 480
| | | | |
QY 818 TCTTCTTGAAGAAGACTATCCAGAGAGCAATGAAGAAGCACTGGCGAGAGGCTTCCAG 877
| | | | |
Db 481 TCTTCTTGAAGAAGACTATCCAGAGAGCAATGAAGAAGCACTGGCGAGAGGCTTCCAG 540
| | | | |
QY 878 CCTCACTCCCGGAGACTTTGGCATTCTCAATGCCGAGAGACCACTTACGGCATGA 937
| | | | |
Db 541 CCTCACTCCCGGAGACTTTGGCATTCTCAATGCCGAGAGACCACTTACGGCATGA 600
| | | | |
QY 938 ATTACTACACATCCGAGCTTCGCGCGCACTAGACGGTCCGTCGCCGAGAGCGACTATC 997
| | | | |
Db 601 ATTACTACACATCCGAGCTTCGCGCGCACTAGACGGTCCGTCGCCGAGAGCGACTATC 660
| | | | |
QY 998 TCGGCGCCATCCATGAGCACGAGAGATTAAGACGCGACGCC 1040
| | | | |
Db 661 TCGGCGCCATCCATGAGCACGAGAGATTAAGACGCGACGCC 703
| | | | |

RESULT 3
CF884351 584 bp mRNA linear EST 31-OCT-2003
LOCUS trico39xel3.b12 T. reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico39xel3, mRNA sequence.
ACCESSION CF884351
VERSION CF884351.1 GI:38139033
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
1 (bases 1 to 584)
Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,
Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
Dean, R.A.
Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
JOURNAL Unpublished (2003)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seg primer: LR-F1 primer.
Location/Qualifiers
FEATURES
source
1..584
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico39xel3"
/dev_stage="mycelial"
/clone_1lb="T.reesei mycelial culture, Version 6 October
2003"
/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match 31.7%; Score 461.2; DB 14; Length 584;
Best Local Similarity 94.0%; Pred. No. 3.3e-81;
Matches 489; Conservative 1; Mismatches 29; Indels 1; Gaps 1;
QY 338 GGGGGGATCAAGCCTTGGGTGACTTTTACCACTGGGATTCCTCCAGCGCTTACG 397
| | | | |
Db 1 GGGGGGATCAAGCCTTGGGTGACTTTTACCACTGGGATTCCTCCAGCGCTTACG 60
| | | | |
QY 398 ATCGCTATGAGGCTGCTCAACGTGGAAGAGTCCAGCTGACTTGGACGGTATGCA 457
| | | | |

Db 61 ATGCTATGAGGCTGCTCAACGTGGAAGAGTCCAGCTGACTTGGACGGTATGCGA 120
| | | | |
QY 458 GGTTCGCTTTGAACCTTTTGGGAGCGAGTCCAGAACTGGATCAATCAAGACCCCT 517
| | | | |
Db 121 GGTTCGCTTTGAACCTTTTGGGAGCGAGTCCAGAACTGGATCAATCAAGACCCCT 180
| | | | |
QY 518 GGAATGAGCCATCTATGATATGACCAAGCGAGCAACCCCGGCGAGAGCAGATTA 577
| | | | |
Db 181 GCAATTAGGCCATCTATGATATGACCAAGCGAGCAACCCCGGCGAGAGCAGATTA 240
| | | | |
QY 578 ACAAGACTTCCACCGAGGCGAACACTGCGCACTGAGCTGCTGCTGGAAGGCCAGA 637
| | | | |
Db 241 ACAAGACTTCCACCGAGGCGAACACTGCGCACTGAGCTGCTGCTGGAAGGCCAGA 300
| | | | |
QY 638 TCATGAGCCATGCGCGCGCGCGCTGCGCTACAGAGGAGACTTGCCTGCAAAAG 697
| | | | |
Db 301 TCATGAGCCATGCGCGCGCGCGCTGCGCTACAGAGGAGACTTGCCTGCAACG 360
| | | | |
QY 698 GCCAGATCCGCACTCTGCTCAACGCGACTATGAGCCTTGGAGCAGCAATGAGCTTC 757
| | | | |
Db 361 GCCAGATCCGCACTCTGCTCAACGCGACTATGAGCCTTGGAGCAGCAATGAGCTTC 420
| | | | |
QY 758 GGGACAAGAGGCTGCTGAGCGACGAGATGGAATTTTCACATTGCTGTTTGGCAATCCCA 816
| | | | |
Db 421 GGGACAAGAGGCTGCTGAGCGACGAGATGGAATTTTCACATTGCTGTTTGGCAATCCCA 480
| | | | |
QY 817 ATCTTCTTGAAGAAGACTATCCAGAGAGCTGATGAAGAC 856
| | | | |
Db 481 ATCTTCTTGAAGAAGACTATCCAGAGAGCTGATGAAGAC 520
| | | | |

RESULT 4
AA415086 1994 bp mRNA linear EST 09-DEC-1999
LOCUS MG0020 RCW Lambda Zap Express Library Magnaporthe grisea cDNA clone
DEFINITION RCW20 similar to Beta-Glucosidase (EC 3.2.1.21), mRNA sequence.
ACCESSION AA415086
VERSION AA415086.1 GI:12537251
KEYWORDS EST.
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Incertae sedis; Magnaporthaceae; Magnaporthe.
1 (bases 1 to 1994)
Wu, S.-C., Bernstein, B.D., Davall, A.G. and Albersheim, P.
Expressed sequence tags of the rice blast fungus grown on rice cell
walls
Unpublished (1997)
JOURNAL Contact: Sheng-Cheng Wu
COMMENT CCRC
University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706 542 4446
Fax: 706 542 4412
Email: wusc@ccrc.uga.edu
Fully sequenced
Insert Length: 1994 Std Error: 0.00.
Location/Qualifiers
FEATURES
source
1..1994
/organism="Magnaporthe grisea"
/mol_type="mRNA"
/strain="CP987"
/db_xref="taxon:148305"
/db_xref="RCW20"
/clone="Mali-1 hermaprodite"
/sex="Mali-1 hermaprodite"
/tissue_type="mycelium"
/dev_stage="Day 5 post-inoculation"
/clone_1lb="RCW Lambda Zap Express Library"
/note="Vector: pBluescript excised from Lambda Zap
Express; Site 1: EcoRI; Site 2: XhoI; Day 5
post-inoculation mRNAs prepared from Magnaporthe grisea
grown at 23C in the dark with constant gyrotory shaking

100 rpm in Vogel's minimal medium containing 0.5% isolated rice cell walls as the sole carbon source. Library provided by Sheng-Cheng Wu. Sequences were processed by one of two methods. Where a full-length alignment to the M. grisea genome sequence was available, the EST sequence was trimmed according to the alignment, otherwise sequence quality was assessed using phred/Phrap version 991019 and trimmed according to phd files (0.05) and for vector segs."

ORIGIN

Query Match 26.6%; Score 387; DB 9; Length 1994;
Best Local Similarity 57.0%; Pred. No. 3.1e-66;
Matches 806; Conservative 1; Mismatches 591; Indels 15; Gaps 5;

17 CTGCGCCCAACGACTTTGATATGGGGCTTCCGACAGCGCCGCTTACCAATGCAAGCCCG 76
290 CTTTCCCAAGGATTTCTTTGGGGCTTCCGACAGCGCTTACCAATGCAAGGCTGCTA 349
77 TCAAGAGAGTGGCCGCGCCGCTTCCATCTGGACAGCTATGCGCACCTGGAGCCATGCG 136
350 TCGACAGAGATGTCGTGGGCGCTTCCATTTGGATATCTTACTGCAATTCGGGTTAAG 409
137 GCACCAAGCGCCCAAGCGCATGTGCTTCCATCACTACCAAGCTTACGATAGAGACT 196
410 TCGCAGAGCGAGCTCGGGTGTGACGCGCATGCGATCACTACCAAGAGAGAGAGACA 469
197 TTGATCTCTGACCAAGTAAGGGGCAAGGCTTACCGCTTCTGTCGTGGTGGGGA 256
470 TTGACCTGCTCAAGTCTGTTGGTCCCAAGTATCCCTTTTCTATCTGGTGGGGA 529
257 TCATTTCCCTCGCGGAGAGGTGATCCGCTCAACGAGAGAGGAATTTGATTTACGCA 316
530 TTATCCCATTTGGTGGCGGTATGACCCCATCAACGAGAGGAGATGACACACTACTCA 589
317 AACTGATTAAGCCCTTGTGAAGGCGGGTATCAAGCCTTGGGTGACTTTGATCACTGG 376
590 AGTTGTAAGAGACTCTCTGAGGCGCGCATCAAGCCACTATCACTCCCTTCCACTGGG 649
377 ATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTCAACGAGAGAGGCTCCAGC 436
650 ATCTGCGGAGATGATTTGACAGAGCGTACGCTGGGTGCTGTAACAGAGAGAGTCCCT 709
437 TGAATTTGAGCGGTATGAGAGGTGTGCTTTGAACGTTTTGGGAGCCGATCCAGAACT 496
710 TGGACTTTGAGCACTATGACAGCGCTCATGTTCAAGGCGCATTTCCAAAGTG---CAAACACT 766
497 GGAATCAACATCAAGACCCCTGATTTAGGCACTATGATATGCAACCGGAGAGCAAG 556
767 GGAATCACTTCAACGAGCCGTTGCTGCTCAATTCGCTTCAAGGCTTGTGCTAGTTTG 826
557 CCCCAGGAGAGAGAGGATTTAAAGAGCTTCAACCGAGAGCAACATGCACTGAGCGCT 616
827 CGCTGCGCGCTGTCTCGACCGGAGAGAGTCTGCTGCGGAGAGCTTCCGAGAGCCT 886
617 GGTCTGCTGGAAGAGCCCAATATGAGCCATGCGCGCGCTGCGCTTACAGAGAGG 676
887 GGAATTTGAGGCGCAACCTTCTGTTGGCCCAAGCTGCTGCTCAAGAGTGTACAGGGAAG 946
677 ACTTTGCGCCCTCGCAAAAGGGCGAGATGGGCACTTCGCTCAACCGGAGATCTATGAGC 736
947 AATTCAAGGCGGAGAGAGAGGAGTGTATGATCAAGTTGAATGGGAGAGCCACTTTC 1006
737 CTTGGAGAGCAATGAGCTTCCGAGACAAAGAGCTGTGAGAGAGGATGGAATTTTACA 796
1007 CTTGGAGATCCCGAAGAGCCCAAGGAGGTGATGACAGCAACCGAGATGATTTCCCA 1066
797 TTGGCTGCTTTGGCAATCCATCTTTTGAAGAAGACTATCCAGAGAGCATGAAGAGC 856
1067 TCTGCTGCTTTGGAGAGCCCATCTACTTTGGCAG---TATCCAGTCTCGATCGTAAAGC 1123
857 AGCTGGGCGAGAGAGCTTCCAGCCTCACTCCGCGGACTTTGCAATCTCAATGCGGAG 916

Db 1124 AGCTTGGCGACCGCTTCCACCTTTTCAAGAGAGAGAGAGAGGCTTTAAGTCAA---GGCTT 1180
Qy 917 AGACCGACTTTTACCGGATGAATTTACTACACATCCAGTTTGGCGCCGCACTTGAAGCTC 976
Db 1181 CAACAGCACTTTTATGGAATGAATGCTTACACGCGCAACTCATCAGGACCAAGAGAGAG 1240
Qy 977 CCGTCCCGGAGAGGACTATCTTGGGCGCATTCATAGACACAGAGAAATTAAGAGCGCA 1036
Db 1241 AGCCGCGCGAGAGAGATTTACTTTGAAACCTTGAACAGTTTCTCAACAGAGCGCGCG 1300
Qy 1037 GCCCGCTTGGCGAGAGAGAGAGCGCGCTCGCTGAGCTGCTTCCCGGACATGTTCCGGA 1096
Db 1301 AATGCAATCGCGCGGAGAGACACATCAACCTGCTTGTCCAAACGACAGAGTTCCGAG 1360
Qy 1097 AGCATCTCGCGCGGTGTACAGCTTGTAGCGCAAGCTC---ATCTACATCACAGAGAGC 1153
Db 1361 AGCTGCTCTGTTTGGCTCAGAGAGCGTTTACCAATCAACCTTGTGTTTACCGAGAAAG 1420
Qy 1154 GATGCCCGTCCCTGGAGAGAGAGATAGACGTTGGAGAGAGCGCGCTCAACGACCCCTTCC 1213
Db 1421 GCACGTGCTCAAGGCGGAGAGAGATATGCGCTGAGAGATCTTCCAGATGACTTCC 1480
Qy 1214 GCATCGGTAATTTGACTCGCACTTGGACTCGATTTTCAAGGCGCATTTACCGAGAGCGG 1273
Db 1481 GGGTCCAGTACTACGACGATTAATGTCAGAGCTTTGGCCAGGCTTATTCGAGAGATGTG 1540
Qy 1274 TCGTCTGCAAGGAGTACTTTTGGTGGCGCTTGTCTCATTAATTTGAAATGTCAGATGCT 1333
Db 1541 TAAACGTCCGCGGTACTCAGCTTGTGCTTATGACAACTTTGATGTCGCGCGAGGCT 1600
Qy 1334 ACAGACCAAGATTTGCGCTCAGCTTCAAGAGACTAC---ACGACCCCTCAAGCGGAGCGCA 1390
Db 1601 ACAGAGACAGAGTTTGGTGTCACTTTGTGCACTACGAGAAACGCGAGAGGCTTACCCA 1660
Qy 1391 AGAAGTCTCCCTGTCTTCAAGAGACATGTTTG 1423
Db 1661 AGAAGAGCGCCAAAGGCAATGAAGCGGTGTTG 1693

RESULT 5
AY109400 2031 bp mRNA linear HTC 17-OCT-2002
LOCUS
DEFINITION Zea mays CL1052.1 mRNA sequence.
ACCESSION AY109400
VERSION AY109400.1 GI:21213111
KEYWORDS
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2031)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
Unpublished (2002)
2 (bases 1 to 2031)
JOURNAL
REFERENCE
AUTHORS
TITLE
COMMENT
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones, these are publicly available in getting from ZmDB and may be found by BLAST searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.
FEATURES
source
1..2031
/organism="Zea mays"
/mol_type="mRNA"

/db_xref="MaizeDB:629921"
/db_xref="taxon:4577"
/clone_jib="Maize Mapping Project/Dupont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed Dupont contigs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

ORIGIN

Query Match 13.8%; Score 201; DB 11; Length 2031;

Best Local Similarity 51.3%; Pred. No. 2,7e-29;

Matches 705; Conservative 1; Mismatches 632; Indels 36; Gaps 9;

55 GCCTACCAAGATCGAAGGCGCGCTCAAAAGAGTGGCGCGCGCGCTTCATCTGGGACACG 114
Db 5 GCGTACCAAGGTTACGAGGGCGCGCTCCACCAACGCGCGCGCGCGCTTCATCTGGGATTTCA 64
Qy 115 TACTGCCACCTGGAGCCATGGCGACCAAGCGCGCGCAATGGCGCTTGGGATCAC 174
Db 65 TTCCGCGACGTCGCCAGGAATATTTGACAGGAATCAAAATGAGACGTTGCGATGATCA 124
Qy 175 TACCACCGCTACGATGAGACTTTGATCTTTGACCAAGTACGCGCGCAAGGCGCTTACCGC 234
Db 125 TACCATGCTACAGGAAGAGTGCATGATCTGATAAAGTTGAACTTTGATGCTTACCGG 184
Qy 235 TTCTCTTGTGTGTGGCGGATCAATTCCTCGCGCGGACGCTGGATCCGTCACGAG 294
Db 185 TTCTCAATCTCATGTGTCAGGATCTCCGCGATGGCGAAGGG-----AAAGTCAATCCA 238
Qy 295 GAGGGAATGATGTTTACAGCAACTGATGACGCCCGTGGAGCGCGGATACAGCGCT 354
Db 239 GAAAGTGAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 298
Qy 355 TGAGTGACTTTGATCACTGCGGATCTGCGTACGCGCTTACGATCGCTATGAGAGCTG 414
Db 299 TACATCAACCTTTACCACTATGATCTTCTTGGCGCTGAGAAATATGAGAGGCTG 358
Qy 415 CTCAACGTGAAAGAGTCCAGCTGGAATTTGAGCGGATGCGAGGTTGTCTTTGAAGT 474
Db 359 TTAGCCGCAAGATGGCGGACTGTG---TTTACAGATGATGCTGACTTGTGTTTAAAGC 415
Qy 475 TTTGGGACCGAGTCCAGAACTGATATCAACCAAGMCCCTGATTCAGGCGCATCTAT 534
Db 416 TACGGCGATGCGTAAAGCACTGATTTAATTAATTAATTAATTAATTAATTAATTAATTA 475
Qy 535 GATATATGCAACCGGACGCAAGCGCCCGGACGAGACGATTAACAAGCACTCCACGAG 594
Db 476 GGTATATGACACAGGCTCAATCTCTCTCAAGGTGACACGATGTG-----CTGCTGT 529
Qy 595 GGCACACTGCTGCACTGAGCCGTGCTGCTGGAAGGCCAGATCATGAGCCATGCCGC 654
Db 530 GGGAAATTCAGCAACGGAACCTTATATGTTGCTCATATTTTCTTGGGACATGCTACT 589
Qy 655 GCGCTGCGCGCTTACAGAGGAGCTTTGCGCCCTCGCAAAAGGCGCGAGATGCGATCTCG 714
Db 590 GCAATGTCAGAGATCCGTACAAAATATCAAGCTCTCAAGAGGTTAAGTGGATGATGTC 649
Qy 715 CTCAACGCGGCACTACTATGAGCCCTGGGACGACATGAGCTCGGAGCAAGAGGCTGCT 774
Db 650 CTGGACTTTCACTGTACGAGGC---TCTTACAACTCACTGATGACCAAGACGACGCC 706
Qy 775 GAGCGACGAGTGAATTTACATTTGCTGTTTGGCAATCCCATCTTTGAGAAGAGAC 834
Db 707 CAAGAGCGCAGGAGCTTCCATTTGCTGTTTGTGATCC---AGCAGAGACGACAC 763
Qy 835 TATCCAGAGACATGAAGAAGACAGCTGGGCGAGAGGCTTCAGGCCCTCACTCCCGCGAC 894
Db 764 TATCCAGAGATGAAGAAGATCTCGTGAAGAGAGGCGCCAGGTTCACTCTGAGAG 823
Qy 895 TTTCATCTCTCAATGCGGAGAGACGCACTTTCAAGGATGAATTAC---TACATCTCC 951

Db 824 GCTAACTGTGAAGGGCTCGGACACTACATCGGTATCAACGAGTACACATCCAGCTAC 883
Qy 952 CAGTTCGCGGCCACCTAGACAGCGGTCCCGTCCCGGAGACGAGATATCTCGGCGCATCAT 1011
Db 884 ATGAAGGGGAGAGAGTGTGTCAGCTGGCGCCAGTACGTAATCTGCGATTGGCAGTT 943
Qy 1012 GAGCACCAGAGATAAGACAGCGAGCGCCGTTGGCGAGAGAGAGCGGCTCGCTGGCTG 1071
Db 944 CAATATGCGATGGCACCGCAATGGCAACCATTTGACACACAGCGAATTTCTAGTGCCTC 1003
Qy 1072 CGCTCTGCGCGGACATGTTCCGAAACATCTCGCCGAGGTATAGCGCTGTACGCGAAG 1131
Db 1004 TACATCGCCCCGACGAGGATGTACGAGGTGGTGAATCACTCAAGAGAGATATGGGAAT 1063
Qy 1132 C---CCATCTACATCAACCGAAGAACGATGCGCCGCTCTGGAGAGAGAACATGACCTGC 1188
Db 1064 CAAAGATCTACATTAAGAGAGAACGAAATGAGACCAAGCTTGA-----AACTTGAACCGA 1117
Qy 1189 GAGAGGCGCTCAACGACGCCCTTCCGATCCGATCTTTGACTCGCACTTGAATCGATT 1248
Db 1118 GACCAATACCTGCGCGACGCGCACGAGAGGTGCGGTTCTACAGAGCTACATCGCGACCTG 1177
Qy 1249 TCCAGGCGATTAACCCAGACGCGCGTGTGTCAGAGGATTAATTTGGTGGCGCTGCTC 1308
Db 1178 AAGAGGCGCATTAAGACGAGGAGCGAAGCTGGC---TGGTACTTTCGCTGCTCTCTCCTC 1234
Qy 1309 GATTAATCTGAATATGATATGCTAGACGACCCAGATTCGCGGCTCACGATTCACAGACTAC 1368
Db 1235 GACAACTTCAGATGGCTGAGGAGGATCTGCTCAAGTTGCGCATCTGCTACGTGGACTTC 1294
Qy 1369 ACCACCTCAAGGCGACGCCCAAGAAAGTGTGCCCTGCTCTCAAGACATGTTT 1422
Db 1295 AACAGCTCGAAGCCACCCGAAAGGCTCGGCTTACTGTTTCAAGGACATGCTT 1348

RESULT 6
CB677158 788 bp mRNA linear EST 09-APR-2003
LOCUS
DEFINITION OSJNBel13N03.f OSJNB Eryza sativa (japonica cultivar-group) cDNA
ACCESSION CB677158
VERSION CB677158.1 GI:29680883
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 788)
Jantasuriyarak,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
JOURNAL
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 13 row: N column: 03
Seq primer: gta aac cga cgg cca gtc.
Location/Qualifiers
1..788
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Niponbare"

FEATURES
source

/db_xref="taxon:39947"
 /clone="OSJNE013N03"
 /tissue_type="leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_1ib="OSJNE"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

ORIGIN

Query Match 13.4%; Score 195.2; DB 14; Length 788;
 Best Local Similarity 56.3%; Pred. No. 2.5e-28;
 Matches 409; Conservative 0; Mismatches 308; Indels 9; Gaps 2;

20 TGCACGAGCACTTTAATGAGGCTTCGCAACGGCGCGCTTACAGATGAGAGCGCGCTCA 79
 60 TCCCGAGAGACTTATCTTCGACCGGCTCGCGCTTATACAGAGAGCGCTTGA 119
 80 AAGAGTGGCCCGCGCGCTTCATCTGAGACAGTACTGCACTGAGCGCATGCGCA 139
 120 ACGAAGCGCGAGAGGCGCGAGCATCTGGGACAGTACATTCAGGCAAAAGTTG 179
 140 CCAAGCGCGCAAGCGCGAGTGGCTTCGATCACTACACCGCTACAGATGAGACTTG 199
 180 AAGATGAGCAACGGAGATGAGAGCTTTTACATGCTCAAGAGAGATCTGA 239
 200 ATCTCTTACCAAGTACCGCGCAAGCGCTTCTCTGTCGTCGTCGAGATCA 259
 240 ACTTGTGACCGGACATGAAATGAGCGCTTCCTTCATTCGTCGAGAGATCC 299
 260 TTCCCTTGGCGGAGGCTGATCCGTCACAGAGAGAGAGATTTTACAGCAAA 319
 300 TGCCAAATGAAACATCAGTGAAGATCAAAAGAGAGATGCTTTTCAACAAGCC 359
 320 TGATTGAGCGCTGTGAGGCGGGTATCAGCGCTGGGTGACTTTGACACAGTGATC 379
 360 TGATTACAGAGTCAATCTAGAGGCTTGAAGCCATTTGTCAACATCTTCAATTCAGCA 419
 380 TGCCTCAGCGCTTACAGATGCTATGAGAGCTGCTCAAGTGAAGAGTCCAGCTG 439
 420 CCCCAAGGCTCTGAGAGCAAAATACGGAGCTTCTCA---GTGAAAACATGTGAAG 476
 440 ACTTTGAGCGTATGCGAGGTTGCTTTGAACGTTTGGGAGCCGAGTCCAGAACTGGA 499
 477 ATTGTGAGCTATGAGCGAGTGTGCTTCGCTGAGATCGGAGCGGGTGAAGTATGGA 536
 500 TCACATCAACGAMCCCTGATTCAGGCTCTATGATATGCAACCGCGCAAGCGCC 559
 537 ACAAGTTCACAGAGCCATGATCTTTCGCGCGCGGCTACGGCTCCGACACCAAGGCC 596
 560 CGGCGAGAGCAG-----CATTAACAAGCACTCCACCGAGGCAACACTGCCACTGAGC 613
 597 CCGGCGGCTGCTCCCGTACGCTTCAAGAAATCGCCCCGGCGACCTCCGCAACAGGC 656
 614 CGTGGCTGCTGGAAGGCCAGATCATGAGCCATGCCCCGCGCTGAGCGCTTACAGCA 673
 657 GCTACGTGCGCGGCAACACTGCTGCTGCGCAAGCGAGCGCTCCGCTTACAGCGCC 716
 674 GGAAGTTCGCGCGCTGCAAAAGGCGAGATCGGATCTGCTCAACGGCGAGTACTATG 733
 717 AAAAGTACAGAGCGAGAGAGAGGCGAGATCGGATCTGCAACAGTGTGCACTGTTCC 776
 734 AGCCCT 739
 777 TGCCCT 782

RESULT 7
 LOCUS CB673495
 DEFINITION OSJNE08A17.1 OSJNE Oryza sativa (japonica cultivar-group) cDNA
 ACCESSION CB673495

VERSION

CB673495.1 GI:29677220

KEYWORDS

EST.

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euharidiales; Oryzaceae; Oryza.

REFERENCE

1 (bases 1 to 786)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,

AUTHORS

Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.

TITLE

Large-scale identification of ESTs involved in the interaction

JOURNAL

between rice and Magnaporthe grisea

COMMENT

Unpublished (2003)
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu

FEATURES

PCR Primers
 FORWARD: gta aac cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 08 row: A column: 17
 Seq primer: gta aac cga cgg cca gtc.
 Location/Qualifiers

source

1. 786
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNE08A17"
 /tissue_type="leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_1ib="OSJNE"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

ORIGIN

Query Match 13.3%; Score 193; DB 14; Length 786;
 Best Local Similarity 57.1%; Pred. No. 6.9e-28;
 Matches 437; Conservative 0; Mismatches 310; Indels 18; Gaps 4;

66 CGAAGCGCGTGAAGAGTGGCGCGCGCTCATCTGGGAGACGATCCACCT 125
 22 CAGGAGCTGTAAAGAGAGACGAGAGGAGACCATCTGGACAGCTTCCGCAAC 81
 126 GAGCCATCGCGACCAACGCGCCCAACGCGATGTGCTTGGATCACTACACCGCTA 185
 82 CTTGGAAGATCAGCGACTTTCAGCAATGCTGATGTGAGTTGATCACTACACCGTTT 141
 186 CAGTAGAGCTTGTATCTTGAACCAAGTACGCGCAAGGCGCTTCTCTCTTTC 245
 142 CGAGAGGATATACACTATGACAGATGAGGATGATGCTATCGCTTCTGATAC 201
 246 GTGTCGCGGATATTCCTCGCGCGGAGGCTGATCCGTCACAGAGAGGAATTGA 305
 202 ATGTGTAAGATCTACCAATGATG-----TTGTCAGTCAATCAAGCTGTATGCA 255
 306 GTTTACAGCAACATGATGAGCGCTGTGAGCGCGGATCAAGCTTGGGTGACTTT 365
 256 CCATCAACAAGAGTATGATGATGATCTTACCAAAAGGAATTCAAGCCATATGTGACA 315
 366 GTACCACTGGAGTCTGCTCAGCGCTTACAGATGCTATGAGAGTGTGCTCAAGTGA 425
 316 CTACCACTGGAGCTTCCCGAGCGCTTGAAGCAAGTCAAGGCGTGGCTTGAAG-- 373
 426 AGAGTCCAGCTGACTTGAACGAGTATGAGAGTGTGCTTGAACGTTTGGGAGACCG 485
 374 -CAGATAGTGAAGATTTGGCGGCGTACCGGAGAGAGTGTCTTCAAGGAGTTGGGAGACG 432

QY 486 AGTCCAGAACTGATCATCAACATGACGACCCCTGATTGAGCCATCTATGATATGACCAC 545
 Db 433 GGTGAAGACATGTGATTCACCTCAACGACGCGCACAGGTGGCCATCCAGGGCTACGACGC 492
 QY 546 CGGAGCAAGACGCCCCGGGAGAGACGACATTA-----ACAGACATCCACCGAGGCAA 559
 Db 493 AGGGCTTCAGAGCCCCGGGCGCTCTCCGTGCTGCTCACCTCTACTGCAAGGCGCGCAA 552
 QY 600 CACTGCACATGAGCCCGGCTGCTGGAAGGCGCCAGATGATGAGCCATCCGCGCGCT 659
 Db 553 CTCGGGACCGAGCCCTACGTCGTTCGCCACCACTTCATCTCGCCACGCGCGCGCGC 612
 QY 660 GGCCTCTACAGCAGGAGCTTTCGCCCTCGCAAAAGGCGCAGATCGGACATCTGCTCAA 719
 Db 613 CACATCTTACAGGACAAATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 672
 QY 720 CGGCACTACTATGAGCCCTGGAGACAGCAATGAGCTCGGAGCAAGAGGCTGTCAGCG 779
 Db 673 CGTATGTGTGATGAGCCGATGTCACAA---CCACATGCAACATCGAGGCGGCAAGAG 729
 QY 780 AGGATGGAATTTACATTTGGCTGTTGCAATCCCATCTTCTT 824
 Db 730 AGCGCAGAGATTTACAGTATGATGTTGCTGATCTTCTTCTT 774

RESULT 8 806 bp mRNA linear EST 08-APR-2003
 LOCUS CB642881
 DEFINITION OSJUNB03B17.f OSJUNB Oryza sativa (japonica cultivar-group) cDNA
 ACCESSION CB642881
 VERSION CB642881.1 GI:29637872

SOURCE EST
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)

KEYWORDS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriaristideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 806)
 Jantschurikar, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe oryzae
 Unpublished (2003)

JOURNAL COMMENT
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: htcp://genome.arizona.edu

PCR Primers
 FORWARD: gta aac cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 03 row: E column: 17
 Seq primer: gta aac cga cgg cca gtc.
 Location/Qualifiers
 1. 806

FEATURES
 source
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJUNB03B17"
 /ribase_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_id="OSJUNB"

ORIGIN
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; 24 hrs after inoculation with Rice Blast (che
 86061)"

Query Match 13.1%; Score 190.4; DB 14; Length 806;
 Best Local Similarity 57.5%; Pred. No. 2,3e-27;
 Matches 407; Conservative 0; Mismatches 286; Indels 15; Gaps 3;

QY 20 TGCCCAACGACTTTGAATGGGGCTTGGCAACGGCCCTTACACAGATGSAAGGCGCGTCA 79
 Db 96 TCCCCAGAGGAGTTCCTTCCTCGGACCGCTCCGCCGACACGATGACGAGGAGCTGTGA 155
 QY 80 AAGAAGTGGCGCGGCGCGCTCCATCTGGGAGACGTAATGCGACCTGAGACCTCGCGCA 139
 Db 156 AGGAGGACGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 215
 QY 140 CCAAGCGCGCAACGCGATGTGGCTTGGCATCTACACCGCTTACATGAGAGACTTTG 199
 Db 216 CGACATTCAGCAATGCTGATGTTGAGATGATGATGATGATGATGATGATGATGATG 275
 QY 200 ATTCTTACCAATGACGCGGCAAGGCGCTTCTCTTCTTCTTCTTCTTCTTCTTCTT 259
 Db 276 AGCTCATGGCAGACATGAGGAGATGATGATGATGATGATGATGATGATGATGATG 335
 QY 260 TTCCTCGGCGGCGGAGCTGATCCGTCACAGGAGGAGGAGGAGGAGGAGGAGGAGG 319
 Db 336 ACCCAATGATG-----TTGATCAAGTCAATCAATCAATCAATCAATCAATCAATCA 389
 QY 320 TGATGACGCGCTTGTGAGGCGGAGGATGACGCTTGGGTGACTTTGATCACTGGGATC 379
 Db 390 TGATGATGACTCTTACGAAAAGAAATTCAGCATATGATGACATCTACATGAGGAGC 449
 QY 380 TGCCTAGGCGCTTACGATCGCTATGAGAGGCTGCTCAACGTGAAAGGTCACCTGG 439
 Db 450 TCCCCAGGCGCTTGAAGCAAGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 506
 QY 440 ACTTGAAGGAGTATGAGAGGTTGCTTGAAGCTTTTGGGAGCGAGTCCAGAACTGA 499
 Db 507 ATTTCCGGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 566
 QY 500 TCACCATCAACGAGCCCTGATTCAGGCACTATGATGATGATGATGATGATGATGATG 559
 Db 567 TCAGCTCAACGAGCGGACGACAGCTGATGATGATGATGATGATGATGATGATGATG 626
 QY 560 CGGCGAGAGGAGC-----ATTACAGACATCCACGAGGCAACATCTGCACTGAGC 613
 Db 627 CCGCGCGCTCTCGGTGCTGCTCCTCACTGCAAGGCGGCACTCGGCAACGAGC 686
 QY 614 CGTGGCTCGCTGGAAGGCGCCAGATGATGATGATGATGATGATGATGATGATGATG 673
 Db 687 CTAAGCTGCTGCGCCACCACTTCATCTGCGCCACGCGCGCGCGCAGATCTACAGG 746
 QY 674 GGGACTTTCGCGCCCTGCAAAAGGCGCAGATGCGCATCTGCTCAACG 721
 Db 747 CAATAATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 794

RESULT 9 889 bp mRNA linear EST 03-JUN-2003
 LOCUS CD438244
 DEFINITION E0JN0511A05.b Endosperm_5 Zea mays cDNA, mRNA sequence.
 ACCESSION CD438244
 VERSION CD438244.1 GI:31353887

KEYWORDS EST.
 Zea mays
 Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
 clade; Panicoidae; Andropogonaceae; Zea.
 1 (bases 1 to 889)
 Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
 Messing, J.

JOURNAL COMMENT
 Sequencing of the maize endosperm ESTs
 Unpublished (2002)
 Contact: Lai, Jinheng
 Dr. Joachim Messing's lab
 Wakeman Institute, Rutgers University

190 Freilinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlaie@akeman.rutgers.edu

Seq primer: T3

FEATURES

source

1. .889
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm_5"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 12.3%; Score 179.2; DB 14; Length 889;
Best Local Similarity 54.8%; Pred. No. 4e-25; Indels 24; Gaps 6;
Matches 493; Conservative 0; Mismatches 383

52 GCCGCTACGAGTCGAGCGCCGCTCAAGAGGTGGCCGCGCCGCTTCATCTGGAC 111
5 GGGGTACGAGTGGAGGGCGCCGCTCAAGAGCGCGCGCTTCATCTGGAGAT 64
112 ACGTACTCCACCTGGAGCCATCGGCAACAGCGCGCAACGCGATGGCTTGGAT 171
65 TCATTGCGGCAAGTCCAGAAATTTGACAGGAATCAAAATGAGAGCTTGCATGAT 124
172 CACTACCAACGCTACGATGAGGACTTTGATCTCTTGAACCAAGTACGCGCAAGGCTTAC 231
125 CAATTCATCGCTTACAGAAAGAGCTCGATCTCATGAAAGTTGAATTTGATGCTTAC 184
232 CGCTTCTCTGTTGCTGTTGCGGATCAATTCCTCGCGCGAGCTGATCCGTCAC 291
185 CGGTTCTCAATCTCATGCTGCGAGATCTCCGAGTGGCGAAGGG-----AAATCAT 238
292 GAGGAGGAATGATTTTACGCAAACTGATGAGCGCTGTGAGGCGGGGTATCAG 351
239 CCAGAGGGTGAAGCGATTAATTAATTTGATTAATCTATCTGCTTACAGAGGCTATG 298
352 CCTTGGGATGACTTTGATCACTGGATGCTGCTCAGCGCTTCAAGATCGCTATGAGAG 411
299 CCTTACATCACTTTACCATATGATCTTCTCTTGGCGCTTGAAGAAATATGAGAGG 358
412 TGGCTCAACGTGAAGAGTCCAGCTGACCTTTGAGCGGTATGCGAGGTGCTTGA 471
359 TGGTAAAGCGCAAGATGCGGAGCTTG---TTTACAGACTATGACTTCTGTTTAA 415
472 CGTTTGGGAGCGGAGTCCAGAACTGGATCACCATCAAGACGACCTGATTCAGGCCATC 531
416 ACCTACGGGATGCGCTTAAGAGCTGTTTACATTTAGAGCAAGATATATAGCGCTA 475
532 TATGATATGCAACCGGAGCAACCGCCCGGCGAGGAGAGCATTAACAGCACTTCACC 591
476 CTGGCTAAGACACAGGTCAATCTCTCAAAAGGTG-----ACGATGCGGTGCT 529
592 GAGGGCAACACTGCCCTGAGCGGTGCTGCTGGAAGGCCAGATCATGAGCATGCG 651
530 GGGGGAATTCAGCAACCGAACTTACATAGTTGCTCAATATTTTCTTGGCACTGCT 589
652 CGGCGGTGGCGCTTACAGCAGGAGCTTGGCCCTCGCAAAAGGGCGAGATCGGATC 711
590 ACTGCACTTGAAGATACCTACGAATATACAGGTGCTCGAAGGATTAAGTGGGAATA 649
712 TCGCTCAACGGCGACTATATGAGCCCTGGAAGAGATGCTCGGAGCAAGAGGCT 771
650 GTCCTGAACTTCACTGATGAGAGGCTTT---ACAACTCACTATGATGACCAAGAGGA 706
772 GCTGAGAGAGGATGGAATTTCACTTGGCTGTTGGCAATCCATCTTCTTGAAGAG 831
707 GCCCAAGAGCCAGGAGCTTCACTTGGCTGTTGTTGATCC---ATTGATTAACGGA 763

QY 832 GACTATCCAGAGCATGAGAGAGAGCTGGCGAGAGGCTTCCAGCCCTCACTCCCGG 891
DB 764 CACTATCCACATATATGAGAGATCTCGAAGAGAGGCTCCAGGTTCACTCCTGAG 823
QY 892 GACTTGGCATCTTCATCCGAGAGACCGACTTCTAGGCAATGATTAATCAATCC 951
DB 824 CA---GGCTAACTGGTGAAGGCTTCGGCAGCACTACATGATTAACAGATCAATCC 880

RESULT 10

CB688082 480 bp mRNA linear EST 04-SRP-2003
LOCUS CEST-23-E-06 Mixed source, strain EP15 and EP15 infected with
DEFINITION hypovirus CHV1-EP13 Cryphonectria parasitica cDNA clone EP15,
EP15-CHV1-EP13 5-prime, mRNA sequence.

ACCESSION CB688082
VERSION CB688082.1 GI:34448951
KEYWORDS EST.
SOURCE Cryphonectria parasitica
ORGANISM Cryphonectria parasitica
Bukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetidae; Diaporthales; Valsaceae; Cryphonectria-Endothia
complex; Cryphonectria.

REFERENCE

1 (bases 1 to 480)
Dawe, A.L., McMain, V.C., Panglao, M., Kashhara, S., Chen, B. and
Nuss, D.L.

An ordered collection of expressed sequences from Cryphonectria
parasitica and evidence of genomic microsynteny with Neurospora
Microbiology 149 (9), 2373-2384 (2003)

TITLE

JOURNAL MEDLINE
PUBMED
22830414
12949163
Contact: Dave AL, Nuss DL
Center for Biosystems Research
University of Maryland Biotechnology Institute
5115 Plant Sciences Building, College Park, MD 20742, USA
Tel: 301-405-7661 or 5111
Fax: 301-314-9075
Email: dawe@umbi.umd.edu
Cryphonectria parasitica EST derived from mixed samples of cDNA
from wild-type (strain EP15) and hypovirus-infected cultures.
Seq primer: M13-reverse.

FEATURES

source

1. .480
/organism="Cryphonectria parasitica"
/mol_type="mRNA"
/strain="EP15"
/db_xref="taxon:38755 and 52571"
/db_xref="taxon:5116"
/clone_lib="Mixed source, strain EP15 and EP15 infected
with hypovirus CHV1-EP13"
/note="Cryphonectria parasitica EST derived from mixed
samples of cDNA from wild-type (strain EP15) and
hypovirus-infected cultures"

ORIGIN

Query Match 12.1%; Score 175.6; DB 14; Length 480;
Best Local Similarity 64.4%; Pred. No. 1.6e-24;
Matches 290; Conservative 0; Mismatches 158; Indels 2; Gaps 2;

QY 11 CGTAGCTTCCCAACACTTTGATGAGGCTTCCGAACGCGCGCTTACAGATGAG 70
DB 4 CGCGTCCGCTACATCAACTTCAAGTGGGCTTTGCGAGCGGCTCTCAAGGTGAG 63
QY 71 GCGCGTCAAGAGAGTGGCGCGCGCGCTCATCTGGGAGACACTACCTGAGC 130
DB 64 GCTTACCGAAGAGAGCGCGCGCGCGCTCATCTGGGAGACACTTCTGAGATCCGG 123
QY 131 CATCGGACCAACGCGCGCGCGCGATGCTTGGCTGATCACTACCGGCTACATG 190
DB 124 GGAATATGCGGAGCGGCTTCAAGCGGCGGCTGCTGCAAGCTACAAACGAGCGG 183

QY	191	AGGACCTTGA	TCCTCTTG	ACAGTA	CGGGCGAA	AGCCCA	CCGCTTCT	CGTGTGT	250
Db	184	AGGACATTG	CGCTGCTC	CAAGAC	CGCTCGGG	CGACCGG	CTTCTG	CTGCTGTGT	243
OY	251	CGCGGATCA	TTCCCTCG	CGCGGAG	CGTGAT	TCCGTC	ACGAGAG	GGAAATTG	310
Db	244	CGCGCATAT	CCCGCTGG	GGCGGCG	CGACGAC	CCCGGTCA	ACAAAGGG	GGCTGAC	303
OY	311	ACAGCAAT	CTGATTGA	CGCCCTGT	TGAGCG	GGGGATCA	CGGCTTG	GGACTTTG	370
Db	304	ACAGCGCTT	CACCAAC	CACTCTG	GGCGGG	GGNATC	AGGCCCTT	CATCAGCTGT	363
OY	371	-ACTGGAT	CTGCCTCA	GGCGCTTCA	CGATTCG	CTATG	AGG-CTG	GCCTCAACGTGG	428
Db	364	AACTGGAG	CACTGG	CGGACG	AGNCTG	GACCGG	CGGTAC	GGGGCTGTG	423
OY	429	GGTCAGCT	GACTTTGA	CGCGGTAT	GGCAG				458
Db	424	GTTCGGCT	CGACTTGT	GGCCGGT	ACGGG				453

RESULT	11		
LOCUS	AY106991	1182 bp	
DEFINITION	Zea mays PCO088410 mRNA sequence.	mRNA	linear
ACCESSION	AY106991		
VERSION	AY106991.1	GI :21210069	
KEYWORDS	HTC.		
SOURCE	Zea mays		
ORGANISM	Zea mays		

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE

1 (bases 1 to 1182)
Hatney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsett,M.S.,
Arthur,L.W., Hanafey,M., Morcane,M., and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 1182)

JOURNAL COMMENT

Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

If you are interested in getting corresponding physical clones, these are publicly available from zmbd and may be found by BLAST searching at MSL.maizemap.org; Zmbd.www.zmbd.iastate.edu; TIGR.www.tigr.org; or NCBI.www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from Zmbd.www.zmbd.iastate.edu.

```

FEATURES
source
location/Qualifiers
1. .1182
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:635002"
/db_xref="taxon:4577"
/clone_idb="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of ESR
assemblies resulting from the application of public
contigs to seed DuPont contig" this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

```

Query Match	12.0%	Score 175;	DB 11;	Length 1182;
Best Local Similarity	54.8%	Pred. No. 3,1e-24;		
Matches 440;	Conservative	0;	Mismatches 345;	Indels 18;
			Gaps	4;

0Y 16 GCTCTGCCCAACGACTTTGATGGGGCTTGGCAACGGCGCCTTACCAGATCGAAGCGCC 75
 |||||

Db	205	GCATTCCCGGAGCGGTTGTCCTTGGGAGCGGCCGGTCGGCGTACCAAGTGCAGGGGATG	264
Oy	76	GTCAAGAAGGTGGCCGCGGCCCGTCATCTGGGACAGTACTGCACTGGAGGCATCG	135
Db	265	GCCAAAGCAGCGCGGGCGGGGCCCCAGCATCTGGGAGCGCTTCATAGAGTTCCGGGACC	324
Oy	136	CGACCAACGGCGCCACGCGAGTGTGGCTTGGCATCACTACACCGCTACGATAGAGAC	195
Db	325	ATCCCTAACAAATGCCACCGCTGACGTGACGGTCCGACGATTCATCGGTACACAGMAAT	384
Oy	196	TTTGAATCTTTGACCAAGTACGGCGCAAGGACCTACCGCTTCTCTTGTGCGTGTCCGG	255
Db	385	GTGAACATAAATGAAGAACATGAGCGCTTGTATGCTGATCCGGTTTTGCATCTCTTGTGTGAGG	444
Oy	256	ATCATCTCCCTCGGCGGCGAGCTGGATCCCGTCAACGAGAGAGGAATTGAGTTTACAGC	315
Db	445	ATTTTCCAGATGGAACTGGCAG-----GTGAACACGAAAGAGTGAATTACTACACAC	498
Oy	316	AAACTGATTTGACGCCCTGTGTGAGGCGGGGTATCAAGCCTTGGGTGACTTTGTACCACTGG	375
Db	499	AGGCTCATAGATTATATGCTCCAGAGAGTATCGGCGCGTATGCAAACTCTTACCATATTAT	558
Oy	376	GATCTGCCCTCAGCGGCTTCAACGATCGCTATGAGAGCGTGGCTCAACGTGAAAGGCTCAG	435
Db	559	GACCTCCCATTTGGCAGCTCATGAAACGTAACCTGGGCTGGCTTAGCC---CAAAAGATTGG	615
Oy	436	CTGCACTTTGACGGGTATGCGAGGTTGTGCTTTTGAACGTTTTGGGAGCCGAGTCCAGAAC	495
Db	616	GAGGGGTTTGGAGACTAGCCGAGTTCGTCTTCCACGCGTTCCGAGAACAGGGTGAAGAAC	675
Oy	496	TGATATCAACATCAACGAMCCCTGGATTACGGCCATATATGATATGCAACGGGACGCAAC	555
Db	676	TGTTTATCTTCAACGAGCCGAGGTCGTGCTGTCTGGGCTACGACAAATGACTTGGAC	735
Oy	556	GCCCCGGGACGAGCAGCATTTAAACAGCACTTCAACCGAGGGCAACACTGCACTGAGCCG	615
Db	736	GCAACGGGAAGGTG-----TTCGGGGTCCCGCCGAGAGCAACTCAACACGAGACCG	789
Oy	616	TGGCTCGCTGAAAGGCCAGATCATGAGCCATGCCCCGCGCGTGGCCGTCTACAGCAGG	675
Db	790	TACCTGTTCGACACATCTCATCTTTCTCATGAGCTGGGGTGAAGCGATPACCGGCAC	849
Oy	676	GACTTTGGCCCTCGCAAAAGGGCCAGATCGGACATCTCGCTCAACGGCGCACTACTATGAG	735
Db	850	AAGTATACGCTTCAACGAAAGGGGAAGATTGAAATTCTCTGGAATTTGTGTGTAACAA	909
Oy	736	CCCTGGGACAGCAATGAGCTTCGGGACAAAGAGGCTGTAGCGACGATGGAATTTTCAAC	795
Db	910	CCTT---TCAGCGACAGCAATGCGGACCAAGGCTCAGACACAGCGACCGAGGACTTTCAC	966
Oy	796	ATTGGCTGTGTTTGGCAATCCCAT	818
Db	967	CTAAGCTGTGTTCTTGAACCCCAT	989

RESULT 12	
CF880605	
LOCUS	CF880605
DEFINITION	tric082xh20.b1.T.reesei mycelial culture, Version 6 October 2003
ACCESSION	CF880605
VERSION	CF880605.1
KEYWORDS	GI:38135287
SOURCE	EST.
ORGANISM	Hypocrea jecorina (anamorph: Trichoderma reesei)
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
	Hypocreales; Hypotracheales; Hypotrachaceae; Hypocrea.
REFERENCE	1 (bases 1 to 662)
AUTHORS	Dienes,S.B., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D., Mitchell,I.T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and Dean,R.A.
TITLE	Analysis of the protein processing and secretion pathways in a Trichoderma reesei EST dataset

Db 496 CTACAGCTCCACTACATCCGCCACCGCAGCTCGCCCGCTCCGCCGACACACCTGTGG 555
Qy 1002 CGCATTCATAGACACAGAGAAATAAGACGGCAGCCCGTGGGAGAGAGCGGCTT 1061
Db 556 CAACGTGAGCTCTTCAACCAACAGAGGAGGATGATCGAGCCCGCAGACGACATC 615
Qy 1062 CGCTGAGCTGCGCTCTGCGCCGACATGTTCGGAAGCATCTGCGCCGCTGTACGCGCT 1121
Db 616 CCCCTGAGCTGCGCCCTGTGCGCCCGCTTCCGGAAGCTTCCTGTGTGATCAGNCAGAG 675
Qy 1122 GTACGGCAA---GCCCATCTACATCACCGGAACGATGCCCGCTTGGAGAGAGAA 1178
Db 676 GTACGGCTACCGCCCATCTACGTGACGGAACGGAACGACGATCAAGGCGAGAGCGA 735
Qy 1179 CATGACGTGAGAGAG 1194
Db 736 CTTGCCCAAGAGAGAG 751

RESULT 14
BE777259 650 bp mRNA linear EST 20-SEP-2000
LOCUS BE777259
DEFINITION MY-26-C-10 PinfestansMY Phycophthora infestans cDNA, mRNA sequence.
ACCESSION BE777259
VERSION BE777259.1 GI:10230914
KEYWORDS EST.
SOURCE Phycophthora infestans (potato late blight agent)
ORGANISM Phycophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
1 (bases 1 to 650)
Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
Initial assessment of gene diversity for the oomycete pathogen
Phycophthora infestans based on expressed sequences
Fungal Genet. Biol. 28 (2), 94-106 (1999)
10587472
10587472
Contact: Govers F
Laboratory of Phytopathology
Wageningen University
Bilthoven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands
Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyco.wau.nl.

FEATURES
source Location/Qualifiers
1..650
/organism="Phycophthora infestans"
/mol_type="mRNA"
/strain="DDR7602, A1 mating type"
/db_xref="taxon:4787"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DHS-alpha"
/clone_idb="PinfestansMY"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Total
RNA was isolated from mycelium of P. infestans DDR7602
cultured for 4 weeks in synthetic medium. EST clones were
named by their position in the microtiter plate, preceded
by the prefix MY (for mycelial) and the successive number
of the microtiter plate (e.g. MY-06-A-04)."

ORIGIN
Query Match 11.7%; Score 169.8; DB 10; Length 650;
Best Local Similarity 58.1%; Pred. No. 2.6e-23;
Matches 319; Conservative 0; Mismatches 224; Indels 6; Gaps 1;

Qy 20 TGGCCACGACTTGTGATGGGCTTCGCAACGGCCGCTACCAATCGAAGCGCGCTCA 79
Db 58 TCCCGGAAGACTTATGTGGGACACGCGCCCTCTGTACCAAGTGGAGGTGCTAGCA 117
Qy 80 AAGAGGTGGCGCGCGCGCTTCATCTGGGACACGTAAGTCCACCTGGAACCATCGCGCA 139
Db 118 ACGAAGTGGACGCGCGACAGCATCTGGAGCGCTTCTCTCGACACCGCGAAGATCG 177

Qy 140 CCAACGGCGCAACGGCGATGTGGCTTGGCATCACTACACCGCTACGATGAGACTTTG 199
Db 178 TGAATGAGACACGGCGCGAGAAAGCCGTGATATATATCTATCCCTTACAGAGAGACTGC 237
Qy 200 ATCTTTGACCAAGTACGGCGCAAGCGCTTACCGCTTCTCTTGTGTGTGCGGATCA 259
Db 238 AACTTATGAAGAAGATGGGACTCAAGCTTACCGTCTTCCATCGCTTGGCCCGGATCA 297
Qy 260 TTCCCTCGCGGCGAGCTGATCCCGTCAACGAGAGGAAATTGATTTACAGCAAC 319
Db 298 TCCCGCAGAGAG-----TCGGCGCAGTGAACGAGAGCGGTGATTCACCAAC 351
Qy 320 TGAATGACCGCTGTGAGGCGGCGGATTCACGCGCTTGGGATCTTGTACCATGGAATC 379
Db 352 TCATCAATGAGCTACTGGCCACGACATACGCGCCCTGGTGAAGCTTACCATTTGGATT 411
Qy 380 TGGCTGAGGCGCTTACGATCGCTATGAGAGCTGAGTCAACGTGGAAGAGTCCAGCTGG 439
Db 412 TACCTTAGCTGTGCAACCGAGTATGACGGTTGGCTGGTGCAAGTATCATCAAGACG 471
Qy 440 ACTTTAGCGGTATGCGAGTGTGCTTTGGAAGCTTTTGGGACCGAGTCCAGAACTGGA 499
Db 472 CCTTCGTGCAATACGCTCGCGTGTGCTTCCAGGCTTCGAGACCGGTGACCAACTGGT 531
Qy 500 TCACATCAACGACGCCCTGATTCAGGCCATCTATGATATGCGACCGGACGACGCC 559
Db 532 TGACGTGAAATGAGCCGTGTGTGCTCCGCTTTTGGGCTATGGAATGATGTTCAATGCT 591
Qy 560 CGGCGACGA 568
Db 592 CAGACGCA 600

RESULT 15
AJ436490 613 bp mRNA linear EST 15-MAR-2002
LOCUS AJ436490
DEFINITION AJ436490 S00007 Hordeum vulgare cDNA clone S0000700006A08F2, mRNA
sequence.
ACCESSION AJ436490
VERSION AJ436490.1 GI:19524942
KEYWORDS EST.
SOURCE Hordeum vulgare
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 613)
Saren,A.-M., Tanskanen,J., Paulin,L. and Schulman,A.H.
Unpublished (2002)
Contact: Schulman AH
Institute of Biotechnology
University of Helsinki
P.O.Box 56 (Valkinkari 6A), University of Helsinki FIN-00014,
Finland.

FEATURES
source Location/Qualifiers
1..613
/organism="Hordeum vulgare"
/mol_type="mRNA"
/db_xref="taxon:4513"
/clone="S0000700006A08F2"
/dev_stage="Shoot"
/clone_idb="S00007"
/note="2-3-4-days after germination"

ORIGIN
Query Match 11.2%; Score 162.6; DB 9; Length 613;
Best Local Similarity 57.2%; Pred. No. 6.7e-22;
Matches 338; Conservative 0; Mismatches 244; Indels 9; Gaps 2;

Qy 143 ACGGCGCAACGGCGATGTGGCTTGGCATCACTACGCGCTTACGATGAGACTTTGATC 202

```
Db      19  ACGGCAGCAAGTGTATACAGCAAGACTTCTATCATCGCTACAAGAGATCTGAAGC 78
QY      203  TCTTGACCAAGTACAGGGCGCAAGGCCCTACCGCTTCTCTTGTGTGTGGCGATCATTC 262
Db      79  TGGTAACTGACATGAATGAGAGCGCTTCGGTTCCTCTTGATGAGAGAGATCCTTC 138
QY      263  CCTCGGCGGAGGCTGGATCCCGTCAACGAGAGAGGAATTGAGTTTACAGCAAACTGA 322
Db      139  CAATGGAACCATGCGCGGAGGATCAACAAGCAGAGATGATTTCTACACAGCCTCA 198
QY      323  TTGACGCCCTGTGAGGCGGGGTATCACGCTTGGGTGACTTTGTACACTGGGATCTGC 382
Db      199  TCGACGAGGTTTGGCTAGAGGGCTGATGCTTTCGTACATGTTCCACTTGCACACC 258
QY      383  CTCAGGCGCTTCAAGATCGCTATGAGGCTGGCTCAACGTGAGAGAGTCCAGCTGACT 442
Db      259  CCGAGGCCCTCGAGAGCAAAATACGGAAGCTTCTTGAGC---GACAACTCGTAAAGGACT 315
QY      443  TTGACCGGTATGCGAGGTTGTGCTTTGAAAGTTTGGGGAACCGACTCGAATCTGATCA 502
Db      316  ACGTGAGTACGCGGAGCTGTGCTTCAAGCTGTTTCGGCGACCGGATGAAATTCTGACCA 375
QY      503  CCATTAACGAMCCCTGGAATTCAGGCCATCTATGATATGACCAGGAGCAACGCCCGG 562
Db      376  CCTTAACGAGCCATGAGTGTCTGCGCTTCGGCTACGGAACCGGACCTTGGCCCCGG 435
QY      563  GCAGGAGC-----AGCATTTAACAGCACTCCACCGAGGCAACACTGCCACTGAGCCGT 616
Db      436  GCCGCTGCTCGCGTACGTCTCCAGAGGCTGGCGGCGCGGGGACTCTCCACGAGGCCCT 495
QY      617  GGCTGCTGGAAGGCCCAATCATGAGCCAGTCCCGGCGCGTGGCCGTTCTACAGCAGGG 676
Db      496  ACATGCGCGGCCAACCTCTCTATGCCCACAGCCGAGGCGGTGCACTGTACCGGACCA 555
QY      677  ACTTTCGCCCTCGCAAAAGGCGCAGATCGGCATCTCGCTCAACGCGGACT 727
Db      556  GGTAACGAGCGCGGCGAGCGGGCCAGATGGCATCGTGCAATATCCCACT 606
```

Search completed: March 28, 2004, 23:31:39
Job time : 3907.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 17:03:34 ; Search time 5752.87 Seconds
(without alignments)
10962.209 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455
Sequence: 1 atgcgcgcagtcgcagctctc.....gggttaagtcgagcgtataa 1455Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl1.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_stb.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_ov.*
22: em_or.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_ay.*
39: em_hlgo_hum.*
40: em_hlgo_mus.*
41: em_hlgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1454.2	99.9	1455	AY281377	AY281377 Hypocrea
2	481.8	33.1	1900	AB003109	AB003109 Humicola
3	452.2	31.1	1695	AY081764	AY081764 Talaromyces
4	397.6	27.3	2171	AB003110	AB003110 Hypocrea
5	373.4	25.7	1590	AY343988	AY343988 Trichoderma
6	278.2	19.1	1828	AK119546	AK119546 Oryza sat
7	278.2	19.1	2032	AY129294	AY129294 Oryza sat
8	253.2	17.4	1988	AF268911	AF268911 Aspergillus
9	248.6	17.1	2045	AK067001	AK067001 Oryza sat
10	247.8	17.0	2068	AF439322	AF439322 Talaromyces
11	246	16.9	1404	AK654334	AK654334 Sequence
12	221.6	15.2	2148	AK067231	AK067231 Oryza sat
13	221.6	15.2	2148	AK120430	AK120430 Oryza sat
14	215.8	14.8	1811	AK066710	AK066710 Oryza sat
15	214.6	14.7	1503	AK652977	AK652977 Sequence
16	213	14.6	1797	AY056828	AY056828 Oryza sat
17	212.2	14.6	302007	SC0939132	SC0939132 Streptomyces
18	210.4	14.5	1689	CSAF26G	CSAF26G Costus spec
19	209	14.4	1665	AK104707	AK104707 Oryza sat
20	205.2	14.1	1792	AK05908	AK05908 Oryza sat
21	203.6	14.0	1792	AK089938	AK089938 Oryza sat
22	202	13.9	1803	AK059210	AK059210 Oryza sat
23	201.8	13.9	1857	AK103027	AK103027 Oryza sat
24	200.2	13.8	1917	AK105026	AK105026 Oryza sat
25	198.6	13.6	1894	AK100165	AK100165 Oryza sat
26	198.6	13.6	1909	OSU28047	OSU28047 Oryza sat
27	196.4	13.5	1842	AK071372	AK071372 Oryza sat
28	194.8	13.4	1824	CF1G13CBET	CF1G13CBET Cellulomonas
29	191	13.1	2081	SSBGL3GB	SSBGL3GB Streptomyces
30	188.6	13.0	298450	SC0939107	SC0939107 Streptomyces
31	186.2	12.8	5996	AF086819	AF086819 Thermobifid
32	183.2	12.6	1796	AK071058	AK071058 Oryza sat
33	182.8	12.6	1734	AK412252	AK412252 Sequence
34	182.8	12.6	1734	AX506587	AX506587 Sequence
35	182.8	12.6	1734	AX651286	AX651286 Sequence
36	182.8	12.6	1881	AF159376	AF159376 Arabidopsis
37	180	12.4	6219	4 OCLPH	X07995 Rabbit mRNA
38	178.6	12.3	1533	AX653834	AX653834 Sequence
39	178.6	12.3	6183	4 OCLPH2	227167 O. cuniculus
40	178.2	12.2	1562	AY142610	AY142610 Arabidopsis
41	178.2	12.2	2016	AX454762	AX454762 Sequence
42	178.2	12.2	2016	AX464120	AX464120 Sequence
43	178.2	12.2	2016	AX491240	AX491240 Sequence
44	178.2	12.2	2016	9 AY358729	AY358729 Homo sapi
45	178.2	12.2	2266	6 AX574385	AX574385 Sequence

ALIGNMENTS

RESULT 1
AY281377
LOCUS AY281377
DEFINITION Hypocrea jecorina strain Qm6a Cel1b (cel1b) mRNA, complete cds.
ACCESSION AY281377
VERSION AY281377.1 GI:317477169
KEYWORDS
SOURCE
ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreales; Hypocrea.
REFERENCE
1 (bases 1 to 1455)
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,

Pred. No. is the number of results predicted by chance to have a

Oliveres, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.
Transcriptional Regulation of Biomass-Degrading Enzymes in the
Flamentous Fungus *Trichoderma reesei*
J. Biol. Chem. (2003) In press
2 (bases 1 to 1455)
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houlek, T.D., Mitchell, T.,
Oliveres, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.
Direct Submission
Submitted (23-APR-2003) Genencor Intl., 925 Page Mill Road, Palo
Alto, CA 94304, USA
Location/Qualifiers
1..1455
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/note="anamorph: *Trichoderma reesei*"
1..1455
/gene="cellb"
1..1455
/gene="cellb"
/note="putative beta-glucosidase"
/codon_start=1
/product="Cellb"
/protein_id="AAP57758.1"
/db_xref="gi:31747170"
/translation="MPSIALPNDPEWGFAPATAVOIRGAVKVGSGPSIMPTCHLP
SRITGANGDVACDHYRDEPDLITTKGAKYFSLSKRIILPLGRLDPVNEGLEP
FYSKLIDALNRGITPWVTLTHMDLPQHLHRYGMLNVEVQLDFEYALCEKRG
DRVQNTITINKPWIQAIYGVATGSNAPERSINRSTEGNTATPEMLGAKQISHR
AVAYSRDPFSPQGOIGISLNGDYEPEDSNEPDRKEAERMEFHIQWPNPILFK
KDYSEKKKQGEPLPALTPADPAFLINAGTFDFGMNYTISOFAHLDGPVPEFDY
AIHEHONKQSGPVGESGLAWLRSCPMPEKHLARVYGLVKGKIDITLTCNCPDGE
NMTEBEAVNDPFRIRYRPSHLSISKAITQGVVYKGFAMALLDNLEMSGYPRRG
VTFDYTYTLKTPKKSALVLMKMPARQVRVAA"

ORIGIN

Query Match 99.9%; Score 1454.2; DB 8; Length 1455;
Best Local Similarity 100.0%; Pred. No. 2,66-262;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCGAGTGGTACTCTGCCCCAAGCATTTGAATGGGGCTTCGCAACGGCGCTTAC 60
Db 1 ATGCCGAGTGGTACTCTGCCCCAAGCATTTGAATGGGGCTTCGCAACGGCGCTTAC 60
QY 61 CAGATCGAAGCGCGCTGCAAGAAAGTGGCGCGCGCGCTTCATCTGGGACACCTACTGC 120
Db 61 CAGATCGAAGCGCGCTGCAAGAAAGTGGCGCGCGCGCTTCATCTGGGACACCTACTGC 120
QY 121 CACCTGAGCCATCGCGACCAACGGCGCAACGGCGCATGTGGCTTGGATCACTACAC 180
Db 121 CACCTGAGCCATCGCGACCAACGGCGCAACGGCGCATGTGGCTTGGATCACTACAC 180
QY 181 CGCTACGATGAGGACTTTGATCTCTTGAACAAAGTACGGCGCAAGAGCTTACCGCTTCC 240
Db 181 CGCTACGATGAGGACTTTGATCTCTTGAACAAAGTACGGCGCAAGAGCTTACCGCTTCC 240
QY 241 TTGTGCTGTGGCGGATCACTTCCCTCGCGCGGACAGGCTGAGATCCCGTCAACGAGAGGGA 300
Db 241 TTGTGCTGTGGCGGATCACTTCCCTCGCGCGGACAGGCTGAGATCCCGTCAACGAGAGGGA 300
QY 301 ATTGAGTTTTCAGCAAACTGATGACGCGCTGTGAGAGCGGGGATATCAACGCTTGGGTG 360
Db 301 ATTGAGTTTTCAGCAAACTGATGACGCGCTGTGAGAGCGGGGATATCAACGCTTGGGTG 360
QY 361 ACTTTGATCAACTGAGATCTGCTCAGGCGCTTACGATCGCTATGAGAGGCTGAGCTCAAC 420
Db 361 ACTTTGATCAACTGAGATCTGCTCAGGCGCTTACGATCGCTATGAGAGGCTGAGCTCAAC 420
QY 421 GTGGAAGAGGCTCCACTGGAATTGAGCGGATATGAGAGGTTGCTTTGGAAGCTTTTGGG 480
Db 421 GTGGAAGAGGCTCCACTGGAATTGAGCGGATATGAGAGGTTGCTTTGGAAGCTTTTGGG 480

QY 481 GACGAGTCCAGAACTGGATACCATCAACGACCCTGGATTGAGGCACTATGATAT 540
Db 481 GACGAGTCCAGAACTGGATACCATCAACGACCCTGGATTGAGGCACTATGATAT 540
QY 541 GCCACCGGACGACACCGCCCGGCGGACGACGACATTTAACAAGACTCCACCGAGGGCAAC 600
Db 541 GCCACCGGACGACACCGCCCGGCGGACGACGACATTTAACAAGACTCCACCGAGGGCAAC 600
QY 601 ACTGCCACTGAGCGGTGCTCGCTGGAAGAGCCCAAGATCATGAGCCATGCCCCGGCGGTG 660
Db 601 ACTGCCACTGAGCGGTGCTCGCTGGAAGAGCCCAAGATCATGAGCCATGCCCCGGCGGTG 660
QY 661 GCCGTTCACAGCAGGGAATTTCGCCCTCGCAAAAGGGCCAGATCGGCATCTGCTCAAC 720
Db 661 GCCGTTCACAGCAGGGAATTTCGCCCTCGCAAAAGGGCCAGATCGGCATCTGCTCAAC 720
QY 721 GCGGACTACTATGAGCCCTTGGGACAGCAATGAGCCTTGGGACAAAGAGGCTGCGAGCGA 780
Db 721 GCGGACTACTATGAGCCCTTGGGACAGCAATGAGCCTTGGGACAAAGAGGCTGCGAGCGA 780
QY 781 CGGATGGAATTTTCACTATGCTGGTGGTTGCCAATCCCATCTTCTTGAAGAAGACTATCCA 840
Db 781 CGGATGGAATTTTCACTATGCTGGTGGTTGCCAATCCCATCTTCTTGAAGAAGACTATCCA 840
QY 841 GAGAGCATGAAAGACAGCTGGGCGAGAGGCTTCACGCTCTCATCTCCGCGGACTTTGCC 900
Db 841 GAGAGCATGAAAGACAGCTGGGCGAGAGGCTTCACGCTCTCATCTCCGCGGACTTTGCC 900
QY 901 ATCTCTCAATGCGGAGAGACCGCATTTTACAGCATGAAATTATCAATATCCAGTTGGCG 960
Db 901 ATCTCTCAATGCGGAGAGACCGCATTTTACAGCATGAAATTATCAATATCCAGTTGGCG 960
QY 961 CGCCACCTAGACGGTCCCGTCCCGGAGACGACTATCTCGGCGCATCTCACTGACACAG 1020
Db 961 CGCCACCTAGACGGTCCCGTCCCGGAGACGACTATCTCGGCGCATCTCACTGACACAG 1020
QY 1021 GAGAAATTAAGACCGGACGCCCCGTTGGCGAGAGAGCGGCTTCGCTGCTGCTCTGTC 1080
Db 1021 GAGAAATTAAGACCGGACGCCCCGTTGGCGAGAGAGCGGCTTCGCTGCTGCTCTGTC 1080
QY 1081 CCGGACATGTTCCGGAAGCATCTCGCCGGGTGTAACGGCTGTACAGGCAAGCCCATCTAC 1140
Db 1081 CCGGACATGTTCCGGAAGCATCTCGCCGGGTGTAACGGCTGTACAGGCAAGCCCATCTAC 1140
QY 1141 ATCACCGAAGACGATGCGGCTGCTGAGAGAGGACATGACGTCGAGAGGCGCTC 1200
Db 1141 ATCACCGAAGACGATGCGGCTGCTGAGAGAGGACATGACGTCGAGAGGCGCTC 1200
QY 1201 AAGCACTTTCGCGATCCGATCTTTGACTCGCATCTGCACTTGAATTCGATTTCCAGGCCATT 1260
Db 1201 AAGCACTTTCGCGATCCGATCTTTGACTCGCATCTGCACTTGAATTTCCAGGCCATT 1260
QY 1261 ACCGAGACGGCGCTGCTGTCACAAAGGGTATCTTGGGGGGGTGTCGATPAACCTTGGAA 1320
Db 1261 ACCGAGACGGCGCTGCTGTCACAAAGGGTATCTTGGGGGGGTGTCGATPAACCTTGGAA 1320
QY 1321 TGCTCAATGAGCTACGAGACCGAGATTGCGGCTCACTGTTCAACAGACTACACCACTTCAAG 1380
Db 1321 TGCTCAATGAGCTACGAGACCGAGATTGCGGCTCACTGTTCAACAGACTACACCACTTCAAG 1380
QY 1381 CGGACGCGCAAGAAATGCTGCTGCTCTCAAGGACATGTTTGGCGCCCGGACAGAGGTT 1440
Db 1381 CGGACGCGCAAGAAATGCTGCTGCTCTCAAGGACATGTTTGGCGCCCGGACAGAGGTT 1440
QY 1441 AAAAGTGGCGGATATA 1455
Db 1441 AAAAGTGGCGGATATA 1455

RESULT 2
AB003109
LOCUS

AB003109 1900 bp DNA linear PLN 25-DEC-2002

DEFINITION	Humicola grisea var. thermoidea bgl4 gene for beta-glucosidase, complete cds.
ACCESSION	AB003109
VERSION	AB003109.1 GI:4249559
KEYWORDS	beta-glucosidase.
SOURCE	Humicola grisea var. thermoidea
ORGANISM	Humicola grisea var. thermoidea
REFERENCE	Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
AUTHORS	1 Takashima, S., Nakamura, A., Hidaka, M., Masaki, H. and Uozumi, T.
TITLE	Molecular cloning and expression of the novel fungal beta-glucosidase genes from Humicola grisea and Trichoderma reesei
JOURNAL	J. Biochem. (1999) In press
AUTHORS	2 (bases 1 to 1900)
TITLE	Uozumi, T.
JOURNAL	Direct Submission
AUTHORS	Submitted (09-MAR-1997) Takeshi Uozumi, The University of Tokyo, Department of Biotechnology, Faculty of Agriculture, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan (E-mail: uozumi@mc.b.t.u-tokyo.ac.jp, Tel: 03-5684-0387, Fax: 03-5684-0387)
FEATURES	Location/Qualifiers
Source	1..1900
gene	/organism="Humicola grisea var. thermoidea"
	/mol_type="genomic DNA"
	/sub_species="IPO9854"
	/db_xref="taxon:5528"
	join(289..332,437..1823)
	/gene="bgl4"
	join(289..332,437..1823)
CDS	join(289..332,437..1823)
	/codon_start=1
	/product="beta-glucosidase"
	/protein_id="BAJ74998.1"
	/db_xref="GI:4249560"
	/translation="MSLPDPFKWGFATPAVOIEGSVNEDEGRPSITWTFCAIPEKIIDGSSGAVCDYKRYKEDIALKELGANSYRFSISWSIIPLEGNDINOGKIDHYK FVDILFAGITPPFIFLHWDLPDALDRYSGFLKEEFADFEVYAIIMPKAIKCKG MITNEPMCSAILGNTGYPAFGRPSRSGPVGDSAREPVIYHNILIHARAVKYY REDKPTGQGEIGITLNGDALTPMDPPDPAIDECDBKIEPAISWPAIDPIYFGKYPSS MRKQDRLPEPTEBEVALYKGSNDPFGAHHYTHYNIKHTGVPEDDPFAGNETLEPY NKYDCCIGPETQSFWLRPHAGCFRDLNMLSKRYGPKIYVTEVETSLKGNDFPLEQ VLEDFRVKXFNDYVRAMAAYVAEDGQNVIRGYLMSLLDNFEMABGYETRFYGVYDY ANDKRRPKXSASAKLPLFSLIRKE"
ORIGIN	
Query Match	33.1%; Score 481.8; DB 8; Length 1900;
Best Local Similarity	61.7%; Pident. No. 3e-80;
Matches	854; Conservative 1; Mismatches 513; Indels 15; Gaps 5;
QY	47 CAACGGCGCGCTTACAGATCGAAGCGCGCTCAAGAAGAGTGCGCGCGCCGCTCAATCT 106
DB	427 CAATCTCAGATTCACGATTCGAGGCGCTCCGTCAACGAGATGCGCGCGCTTCATCT 486
QY	107 GGACACGATGCGCACCTTGAGCGCATCGCGACCAACGGCGCGCAACCGCGATGTGCTT 166
DB	487 GGGACACCTTCGCGGCATCCCGGCAAGATCGCGACGCACTCGGGCGCGTGCGCT 546
QY	167 GCGATCACTACCAACCGCTAGATAGAGACTTGATCTTTTACCAAGTAACGCGCGCAAG 226
DB	547 GCGACTCGTACAAACGCAACGAGAGGACATTCCTCTTAAGAGAGCTCGCGCGCAACT 606
QY	227 CCTACCGCTTCTCTTGTCGTGTGCGCGATCATTTCCCTCGGCGCGGACGTGATCCCG 286
DB	607 CGTACCGCTTCTTCATCTCGTGTGTGCGCAATCATCCCTCGGGCGCGCAATACCCCA 666
QY	287 TCAACGAGAGGGAATTGATTTTACAGCAAACTGATTTAGCGCCCTGTGAGCGGGGTA 346
DB	667 TCAACGAGAGGAGATCGACACCACTACGTCAAGTTTGTTCAGAGACTGTACAGCGCG 726
QY	347 TCAGCGCTTGAGTGACTTTGACACAGTGGGATTCGCGTCAAGCGCTTCAGATGCGATG 406
DB	727 TCAGCGCTTATCACTCTTTTCCACGTGGGACCTGCGAGCGCCCTCGACCAAGCGCTTAC 786

ORIGIN

Query Match	33.1%	Score 481.8	DB 8	Length 1900
Best Local Similarity	61.7%	Pred. No. 3e-80		
Matches 854		Conservative 1	Mismatches 513	Indels 15
				Gaps 5

OY	47	AAAGGGCGCGCTTACAGATCGAAGGCGCGGCAAGAAAGGTGGCGCGCCGTCATCT	106
Db	427	CAACTCACAGTACCAAGATCGAAGGCTTCGTCACAGAAATGGCCGTGGCCGTCATCT	486
OY	107	GGGACAGTACGTGCACCTTGAGGCATCGCCCAACGCGCGCAACGGCATGTGGCTT	166
Db	487	GGGACACTTGTGGGCCATTCGCCGCAAGATCGCCGAGGACGTGGGGCCCGTGGCTT	546
OY	167	GGCATCACTACACCGCTTACGATGAGCATTTGATCTCTTTCAGCAAGTACGCGCAAGG	226
Db	547	GGCATCTGTACAGGGGACCAAGGAGGCAATGGCCCTCTCAAGAGAGTGGCGCCAACT	606
OY	227	CTTACCGCTTCTCCTTGTGTGTGTGGGATCATTTCCCTCGGGGCGACGCTGATCCCG	286
Db	607	CGTACCGCTTCTCATCTCGTGTGTGCGCATCATCTCCCTCGGGCGCGCATGACCCCA	666
OY	287	TCAAACGAGGAGGAATTGATTTTACAGCAAACTGATTGAAGCCCTGTGAGGGGGGTA	346
Db	667	TCAAACCAAGAGGAGTCGACCACTACGTCAAGTTGTGAGAGACTGTATCGAGGCGCGCA	726
OY	347	TCAGCGCTTGGGTGACTTTGTACCACTGGGATTCGCCACAGCGCTTACAGATGCGATG	406
Db	727	TCACCCCTTATCAACCTCTTTCACATGAGGACTTGGCCGACGCCCTCGACAAAGGCTTACG	786

QY	407	GAGCCTGGCTCAACG	GTGAAGAGGCTCCAGCTG	GACTTTGAGCGGTAATG	CCAGAGTTG	AGCT	466										
Db																	
Db	787	GGCGCTCTCTCA	CAGAGAGAGT	TGCCCGCCGAC	CTTTGAGAACTACG	CCCCGATCATGT	846										
QY	467	TTGAA	CGTTTGGGAG	CCGAGTC	CAAGATCGAT	CACATCAAGAMCCT	TGACG 526										
Db																	
Db	847	TCAAGG	CATTC	CCCAATG	---CAAG	CACTGATATCACTTCAACGAG	CCCTGGTCCG 903										
QY	527	CCATCTAT	TGGA	TATG	CCACCGGAG	CAAGCCCGCCGAGACAG	ATTAACAAGACT 586										
Db																	
Db	904	CCATCTCTGG	GTACAA	ACGAGCT	ATCTTGC	CCCCCGGCA	CACCTCGAGCCGACG	CAAGT 963									
QY	587	CCACCGAGG	GCACAC	TGACCTGAG	CCGTGAGT	AAAGGCC	CACATATGAGCC 646										
Db																	
Db	964	CCCCCGT	CGGAG	CAGGCGCCCGGAG	CCCTGAT	TGTGGGAC	CAACATCTCTATG	CGCC 1023									
QY	647	ATGCCGCG	CGCGT	TGAGCCCTT	CACAGAG	CACTTGC	CGCCCTTGCAAAAGG	CGCAGATCG 706									
Db																	
Db	1024	ACGCGCCG	CGCGT	TCMA	GGCTTAC	CGCAAG	CACTTCAAGCC	CACCCAGGCGGAG	ATCG 1083								
QY	707	GCATCTGG	CTCAAC	GGGCG	CTACTAT	GAGCCCTG	GGGAGCAGCAAT	TGAGCTTC	GGGAGCAAG 766								
Db																	
Db	1084	GCATCAC	CCCTTG	AGACGGG	AGCGCAC	CCTGCCCTT	GGAGCCCGG	AGACCCG	CGCACATMG 1143								
QY	767	AGGCTGT	GTAG	CCGACG	AGATTTT	CAATYGG	CTGGTTG	CCCAATCC	CATCTTCTTGA 826								
Db																	
Db	1144	AGGCTGT	GCAG	CCGAA	GATCG	AGTTG	CGC	CAATCTGT	GTGTTGCGCGAC	CCCATCTATCTTG 1203							
QY	827	AGAGAG	CTATCC	CAGAG	AGCAT	TAAGAA	AGCAGCT	GGGCGAG	AGAGAGCTTCC	AGCCCTCACTC 886							
Db																	
Db	1204	GCAAGT	AC---CCG	CACTCG	ATG	GGCAG	AGAGAG	AGCTGGG	CGACCCGCTG	CCGAGTTCACTC 1260							
QY	887	CCGCGAG	CTTGTG	CCATCT	CTCAAT	TCGCGAG	AGACCG	CACTTCTA	CGG	CATGAATTTACTACA 946							
Db																	
Db	1261	CCGAGAG	GTGCG	CCCTG	GTCA---GGG	CTG	CAACG	ACTTCTA	CGGATG	AGCACTACA 1317							
QY	947	CATCCAG	TTTGCG	CGCCAC	CTTGA	CGGTCCG	TCCCG	AGACG	AGATAT	TCTGGGCCCA 1006							
Db																	
Db	1318	CGGCA	ACTAT	CA	TCAC	AGACAG	CGGCG	CTG	CGCCCG	AGAGACG	ACTTCTCGGCAAC 1377						
QY	1007	TCGAT	TGAG	CA	CCAG	AATTA	AGGAG	CGGCG	CCGTT	GGCGAGAGAG	CGGCGCTCGCT 1066						
Db																	
Db	1378	TCGAG	AG	CTCTTT	TAC	ACAG	ATGACG	GCAC	ATG	CGATCG	CGGAGACCA	CACTTCTTCT 1437					
QY	1067	GGCTG	CGCTC	TG	CCG	AGAT	GTTCG	GAAG	ACAT	TCGCG	CGG	GTGATCGG	CGCTGTAGC 1126				
Db																	
Db	1438	GGCTG	CGCG	CGCA	CGCC	CAAG	GGCTTCCG	CGAC	CTG	CAACTG	GGCTC	AGCAAG	GGCTAGC 1497				
QY	1127	GCAAG	CCCC---AT	TTC	AT	CAT	CAC	CGAG	AACG	AGATGCC	GTGCC	CTG	AGAGAGAG	AACTGA 1183			
Db																	
Db	1498	GCTAC	CCCCCA	AGAT	TTC	AGT	ACCG	AAAG	CGAC	CTTGC	CTC	AGAG	GGGAG	AAAC	CAACATGC 1557		
QY	1184	CGT	CGAG	AG	AG	CG	CGT	CAAC	CGAC	CCCTT	CGCAT	CCG	STACTT	TGACTCC	CACTTGGACT 1243		
Db																	
Db	1558	CCCT	CGAG	AG	AG	GT	CTCTG	AGAG	AC	ACTTCC	CGCG	CTCA	AGTACTT	CAAC	ACTACGTGTGGCG 1617		
QY	1244	CGAT	TTTCA	AGG	CCATTA	CC	AGAG	CGG	CGT	CGT	CTCA	AGG	GGTACTT	TTC	CGTGGCGT 1303		
Db																	
Db	1618	CCAT	TGG	CGG	CGCG	CGT	CGCG	AGAG	CGG	CTG	CAAC	CGT	CCCG	GTTC	CTG	CGCCGTG	CGC 1677
QY	1304	TGCT	CGAT	TA	CTT	TGGA	TGG	CTA	CGG	AC	CCAG	ATTC	CGG	CGT	CA	CGTTC	CAG 1363
Db																	
D																	

Dib

```
RESULT 3
AY081764      1695 bp  mRNA  linear  PLN 31-MAY-2002
LOCUS       Talaromyces emersonii beta-glucosidase (bg11) mRNA, complete cds.
DEFINITION  AY081764
ACCESSION   AY081764.2 GI:21284364
VERSION     AY081764.2
KEYWORDS    Talaromyces emersonii
SOURCE      Talaromyces emersonii
ORGANISM    Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; Talaromyces.
REFERENCE   1 (bases 1 to 1695)
AUTHORS     Murray, P.G., Collins, C.M. and Tuohy, M.G.
TITLE       The molecular cloning of the beta-glucosidase genes of Talaromyces
            emersonii
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1695)
AUTHORS     Murray, P.G., Collins, C.M. and Tuohy, M.G.
TITLE       Direct Submission
JOURNAL     Submitted (04-MAR-2002) Biochemistry, National University of
            Ireland, Galway, Galway, Ireland
            3 (bases 1 to 1695)
REFERENCE   Murray, P.G., Collins, C.M. and Tuohy, M.G.
AUTHORS     Direct Submission
JOURNAL     Submitted (31-MAY-2002) Biochemistry, National University of
            Ireland, Galway, Galway, Ireland
REMARK      Sequence update by submitter
COMMENT      On May 31, 2002 this sequence version replaced gi:19568099.
FEATURES
            location/Qualifiers
            source
            1..1695
               /organism="Talaromyces emersonii"
               /mol_type="mRNA"
               /db_xref="taxon:68825"
            1..1695
               /gene="bg11"
               /EC_number="3.2.1.21"
               /note="family 1 glycosyl hydrolase"
               /codon_start=1
               /product="beta-glucosidase"
               /protein_id="AA09551.2"
               /db_xref="GI:21284365"
               /translation="MTTETPTTTTSTLSPDPLMGFNATSYOIEGAFDEDEGRLPSIM
               DTSKTPGVDEGTNGDVACDSYHRTBEDILKKGAKYRSLSPRIIPGGRND
               PINEKQGFYSKPLDLDLAAGIEPFTLFFMDLPDELMKRYGKLPKEEVAAYANA
               RVFNALGSKVKHWITENBPMGSGVLGHTGKXAPGRTSPRTSGPDEGVRDEPWIG
               NLVAAGTVVDIYAREFEKOGELGTLGDNAEPWDPENPDVRCAPDKIEFALSM
               PADPIYHGXKPDMSVKQIDRLPKPTBEHAFYHGSNDIFYGMNHYCENTYRNRTGEPD
               PEDLAGNDLIMEDKGNPFGPETQCEMLRPFPLGFKLTKMLADRYNNPKIYVTEG
               TSVAGSDKPLEEVLNDEFVQYRYDIIGAMVDVADGVNVAYAMVMSLLDNFEMSE
               GYSRFRFTVYVDYKNGQKRIPKSALVIGELFNKYIRKE"
ORIGIN
Query Match      31.1%; Score 452.2; DB 8; Length 1695;
Best Local Similarity 59.5%; Pred. No. 1.1e-74;
Matches 833; Conservative 1; Mismatches 559; Indels 12; Gaps 4;

QY 17 CTCTGCCAAGACTTGAATGGGGCTTCGCAACGGCCGCTACAGATGGAAGCGCCG 76
Db 147 CTCTTCTCCGACTTCCTCGGGGTTTCGCACTGCACTGCACTGCACTGCACTGCACTG 206
QY 77 TCMAAGAAGTGGCGCGCCGCTTCATCTGGGACACGTACTGCACTGGAACCATCGC 136
Db 207 TCAGCAGGATGGCGCGCTTCCTGATATGGGACACTTTCACAGACGCCCGGCAAGG 266
QY 137 GCACCAACGCGCGCAACGGCGATGGCTTGGCATCTACCAACCGCTAGATGAAGACT 196
Db 267 TCAGAGACGGGACCAATGGCGATGTCCCTGCGACTGTAACAACGATCCGATGAAGACA 326
QY 197 TTGATCTCTTGACCAAGTACGGCGCAAGGCTACCGCTTCTCTGTCGTGGTCGCGA 256
Db 327 TTGAGATATTGAAGAATACGGTGCACCAAGATATGCTTCTGTTGCTGCGCAAGTA 386
```

```
QY 257 TCATCCCTCGCGGCAAGCGTGATCCGTCACAGAGAGGAATTGATTTACAGCA 316
Db 387 TCATCCCTCGCGGCGGCGGAAGCAAGCCCATCATGAAAAGGCGCTCATCTTACATCCA 446
QY 317 AACTGATGAGCGCCCTTGAGGCGGTATCAACGCTTGGGTGACTTTGTACACTGGG 376
Db 447 AGTTCTTGATGATTTACAGCCGCGGCGATCAGACCTTGTGTACGCTTCCACTGGG 506
QY 377 ATTCGCTCAGCGCTTCAAGATCGCTATGAGAGCTGGCTCAACGTGAGAGAGTCCAGC 436
Db 507 ATTCGCCGACGAGCTGATGAAAGCGGTAGCGGGGCAAGCTCAACAGAGAGAGTTGTGG 566
QY 437 TGAATCTTGAGCGGTATGCAAGGTGTGCTTTGAACGTTTGGGAGCCGATCCAGACT 496
Db 567 CCGACTAGCCCACTATGATCCGCTGTCTTCAACGCGCTCGGCTGMAAGTCAAGCACT 626
QY 497 GATACACATCAACGAMCCCTGATTCAGGCACTTATGATATGCAACCGGAGCAAG 556
Db 627 GATCAGCTTTCAGAGACCTTGATGCAATTCGCTGCTCGGCGCAACACCGGAGACAG 686
QY 557 CCGCGGCGAGAGCAGCATTTACAGCACTCCAGAGGCAACACTGCACTGACCGCT 616
Db 687 CCGCGGCGGAGAGAGGATCGGACGAAAGTCCGAGGAGACGGCACTCGGAACTT 746
QY 617 GCGTCGCTGGAAGCGCCAGATCATGAGCATAGCCCGCGCTGCGCTTACAGCAGG 676
Db 747 GATTCGAGGTCAAACTTCTGTGCTCATATGTACCGTTTGATCATCTTATCGAGAG 806
QY 677 ACTTCGCGCCCTCGCAAAAGGCGCAGATGCGGCACTCTCGCTCAACGCGCACTATAGC 736
Db 807 AGTTCAAGAGAAACAAGGCGGGAATGGATATCACTTCAACGCGCACTGGCCGAAC 866
QY 737 CCTGGGACAGCAATGAGCTTCGGGACAAAGAGGCTGCTGAGCGACGATGAATTTACA 796
Db 867 CATGGATCCCGAAACCCAGACAGATCGAAGCTCGCAACGAAAGATCGAATTCGCA 926
QY 797 TTGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGCACTATCCAGAGACATGAANAAC 856
Db 927 TCTCGTGTGTTGCGGAGCCGATTTACAGGGAATAC--CGGACACATGTTGTAAC 983
QY 857 AGCTGGCGAGAGAGCTTCAGCCCTTCACTCCGCGGACTTTCGATCCATCGCGGAG 916
Db 984 AGTTGGCGCACCGTTCGCGGAAGTTCAAGCCGGAAGAAATCGCTTCGTCGCA--CGGCA 1040
QY 917 AGACGCACTTTCACGCGCAAGATTTACTACACATCCAGTTCGCGCGCCACTAGACGTC 976
Db 1041 GCACGATTTTACGCGCATGAACCACTACTGCGAAACCTACATCGCAACCGGACCGGCG 1100
QY 977 CCGTCCCGAGACGCACTATCTCGCGCCATCATGAGCACAGAGAGAAATGAAGCGCA 1036
Db 1101 AACCAAGACCGGAGGACATGCTGGGACATCTGAGATCTGTATGAGAGCAAGAACGCA 1160
QY 1037 GCGCCGTTGGCGAGAGAGCGGCTCGCGCTGCGCTCTCTCCCGGACATGTTCCGA 1096
Db 1161 ACCCATTCGCGCCCGAGACCAATGCGAGTGGCTCGCGCGTTCGCGGAGTTCAAGGA 1220
QY 1097 AGCATCTGCGCGGCGGTATCGGCTGTACGGAAGCCC--ATTCACATCAACGGAAGC 1153
Db 1221 AGCTGCTGAAGTGGCTTGGAGACCGGTATACAAACCCCAAGATTCACGTAACGGAACG 1280
QY 1154 GATGCCCGTGCCTGGAAGAGAAACATGACGTGCGAGAGGCGCTCAACGACCCCTTCC 1213
Db 1281 GGAACGAGCGTCAAGGCGGAGAGCAAGACCGCTCGAGGAAGTGTCAACAGATGAGTTCC 1340
QY 1214 GATCGGATATCTTGAATCTGCACTTGAAGTCAATTTCAAGGCCATTTACCAAGAGCGCG 1273
Db 1341 GGTGCAAGTACACGAGGACTATATACGCGGCGAGTGTGCTGTCACAGAGATGAGGAG 1400
QY 1274 TCGTGTCAAGGGGATCTTGGGTGGCGTGTGCTGATTAACGTGGAATGATGATGCT 1333
Db 1401 TCAAGGTCAAGGCTTACATGCTTGGAGTTGTTAGACAACTTTCGAATGATCCGAAGCT 1460
```


Qy 1299 GGCCTGCTCGATACTTGATGATGCTAGATGCTACGGAACCGAATTCCGCTCAGCTT 1358
Db 1879 GTGCTCATGGAACAATTGAGTGGGGCGGACGGCTAGTGAAGAGTTTGGGGTTACGTA 1938
Qy 1359 CACGACT--ACACCACTTCAGACGCGCCCAAGAACTCTGCTGCTTCAAGGA 1415
Db 1939 TGTGATTATAGATAGGCGAAGCGGTTCCCAAGAAAGCGCAAGAGCTTGAAGCC 1998
Qy 1416 CATGTTTG 1423
Db 1999 GCTGTTTG 2006

RESULT 5
AY343988 1590 bp mRNA linear PLN 13-AUG-2003
LOCUS Trichoderma viride strain AS 3.3711 beta-glucosidase 2 (bg12) mRNA,
DEFINITION complete cds.
ACCESSION AY343988
VERSION AY343988.1 GI:33521681
KEYWORDS
SOURCE Trichoderma viride
ORGANISM Trichoderma viride
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; mitosporic Hypocreales;
Trichoderma.
REFERENCE 1 (bases 1 to 1590)
AUTHORS Liu,B.D., Yang,Q. and Zhou,Q.
TITLE Cloning and Sequence Analysis of the beta-glucosidase 2 Gene from
JOURNAL Trichoderma viride AS 3.3711
REFERENCE 2 (bases 1 to 1590)
AUTHORS Liu,B.D., Yang,Q. and Zhou,Q.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2003) Department of Life Science and Engineering,
Harbin Institute of Technology, Harbin, Heilongjiang 150001, P. R.
China

FEATURES
source
1..1590
/location="Qualifiers
/organism="Trichoderma viride"
/mol_type="mRNA"
/strain="AS 3.3711"
/db_xref="taxon:5547"
1..1590
/gene="bg12"
192..1544
/gene="bg12"
/product="beta-glucosidase 2"
/codon_start=1
/protein_id="AAQ21384.1"
/translation="MLPKDPMWGRATAYOIEGAVDDGRPSIWMPCAPQPKIADG
SSGVTACDSYKRTMEDIALKSLGAKSIRSEISRSRIIPBGRGDANVQADHYKRV
VNDLLDAGITPITLFTFMDLPEGLHQRGGLNTERPLDEENARVAFRLPVRW
ITFNEPLCSAIPGSGSFAPRQSTSEBPTWGHNIIVAHGRAVAVADPFKPSGDG
QIGVLNGDFTYPMDAADPADKAEARLEFFTAADPIYLGDPASMRKQLDRLP
TFPEERLIVHSGNDPFGMNYHTSNYIRHSSPASTPSATSCSSPSRATASAPR
SPRLAPYRPAATSWCGSARGTPIYVTENSTIKGESDLPKELILEDPRVRYKYNRY
IRAVTVAVELDGVNKGTFAMSLMDFEMADGVTRFGVTVVDYENQKSPRAQRA
"

ORIGIN
Query Match 25.7%; Score 373.4; DB 8; Length 1590;
Best Local Similarity 59.4%; Pred. No. 6.2e-60;
Matches 802; Conservative 1; Mismatches 497; Indels 51; Gaps 8;

Qy 20 TGCCCAAGCACTTGAATGAGGCTTTCGCAAGCGCGCTTACAGATCGAAGCGCCCTCA 79
Db 196 TGCCCAAGCACTTTCAGTGGGGTTTCGCAAGCGCGCTTACAGATCGAAGCGCCCTCG 255
Qy 80 AAGAAGTGCGCGCGCGCGCTTCATCTGGACAAGTACTGCCACCTGGAGCCATCGCGCA 139
Db 256 ACCAGACGCGCGCGCGCGCGCTTCGCGACAGCTCTGCGCGCAGCCCGCAAGATCG 315

Qy 140 CCAACGCGCGCAACGCGATGATGCTTGCATCACTACACCGCTACGATGAGGACTTTG 199
Db 316 CCGAGCGCTGTCGGGGGTGTAACGCGCTGCGCATCTGTACAAACCGACGCGGAGCACTTG 375
Qy 200 ATCTTTGACCAAGTACGCGCGCAAGGCTTACCGCTTCTCTTGTGCTGCTCGGATCA 259
Db 376 CGCTGCTCAAGTGGCTGGGGCCAAAGACTACCGCTTCTCAATCTCGTGTCCGCACTCA 435
Qy 260 TTCCCTCGGGGCGAGCGCTGATCCGTCACGAGGAGGGAATTGATTTAAGCAAAC 319
Db 436 TCCCGAGGCGGCGCGCGATGCGTCACACGAGGCGCATGACCACTAGCTCAAGT 495
Qy 320 TGAATTGACGCTTGTGAGCGGGGTATCAACGCTTGGGTGACTTTGTAACACTGGGATC 379
Db 496 TCGTCACACGACTGTCTGACGCGCGCATCAACGCTTTCATCAACCTTTTCACTGGGACC 555
Qy 380 TGCTCAGGCGCTTACGATCGCTATGAGAGCTGGCTCAACGTGAAAGAGTCCAGCTGG 439
Db 556 TGCCGAGGGCTGTGATCAGCGGTACGCGGGGCTGTGAACCCGACCGAGTTCCCGCTCG 615
Qy 440 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGGACCGAGTCCAGAACTGA 499
Db 616 ACTTTGAAACTACGCGCGGCTCATGTTCAGGGCGCT--GCCCAAGGTGCGCAACTGA 672
Qy 500 TCACATCAACAGAMCCCTGATTCAAGGCATCTATGATATGCAACCGGAGCAACGCC 559
Db 673 TCACCTTCAACAGACCGCTGTGCTCGGCACTCCGGCTACAGGCTCCGGACCTTGCGCC 732
Qy 560 CGGCGAGGACAGCAATTACACAGCACTCCACGAGGCAACACTGCCACTGAGCGCTGGC 619
Db 733 CCGGCGCGGACAGCA-----CCTCGAGACCGTGGGA 762
Qy 620 TCGCTGAAAGGCCCGCATGATGAGCCATCCGCGCGCGTGGCGCTACAGAGGAGCT 679
Db 763 CCGTGGGCAACACATCTGTGCGCCACGCGCGCGCTGCAAGCGTACCGGACACT 822
Qy 680 T---TCGCCCTCGCAAAAGGCGAGATCGGCACTTGTGCTCAACGCGGCACTACTATAGC 736
Db 823 TCAAGCCCGCAACGCGGAGCGGCCAGATCGGCATCTGCTCAACGCGGACTTCACTAAC 882
Qy 737 CTTGGACAGCAATGAGCTTCTGGGACAGAGGCTGTGAGCGACGATGGAATTTACA 796
Db 883 CTTGGACAGCGCGCGACCGGCGGACAGAGGCGCGGCGCTGAGTCTTCA 942
Qy 797 TTGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTTCCAGAGAGATGAAGAGC 856
Db 943 CGGCTGTTTGGGATCCATCTACTT--GGGCGACTACCGGCGTCAATGCGCAAGC 999
Qy 857 AGCTGGGAGAGAGCTTCCAGCCCTCACTCCGCGGACTTTGGCATCTCAATGCGGAG 916
Db 1000 AGCTGGGCGACGCGCTCCGACCTTAAAGCCGAGAGCGCGCTCTGCTCA---CGGCT 1056
Qy 917 AGACCGACTTCTACGCGATAATTACTACACATCCAGTTCGCGCGCCACTTGAACGCTC 976
Db 1057 CCAAGCACTTTTACGCGATGAACCACTACAGTTCACATCCGCAACCGGACGCTCGC 1116
Qy 977 CCGTCCCGAGAGAGATATCTCGGCGCCATCCATAGGACACGAGAAATAAGAGCGGCA 1036
Db 1117 CCGCTC---GACGACACCGTGGCAAGCTGACGCTTTCACCAACAGAGAGGCA 1172
Qy 1037 GCCCGTTTGGAGAGAGGCGCGCTGCTGCTGCGCTCTGCGCGGACATGTTCCGGA 1096
Db 1173 ACTGATCGGCGCGCGGAG--CGAGTCCCGCTTGGCGCTGTGCGCTCCCGCACTTC 1231
Qy 1097 AGCATCTCGCGCGGTGTACGCGCTGTACGCGCAAGCCATCTACATCAGGAAACGAT 1156
Db 1232 CTGGTGTGATCAAGCAAGGTAC---CTACCGGCCATCTAGTACGGAACGCGCA 1287
Qy 1157 GCCCGTCCCTGAGAGAGAAACATGACGTGCGAGAGGCGCGTCAACGACCCCTTCCGCA 1216
Db 1288 CGAGCATCAAGGCGAGAGGACTTGGCCAAAGAGAAATTTCTGAAAGATGACTTCAGGG 1347

Oy	1217	TCCGATCTTTGACGTGGCAGCTTGACACTGCATGATTCCAAAGGCCATTACCAGGACCGCGTCG	12176
Dd	1348	TCAAGTACTTAACAACGATCAATCCGTCCATGGTTAACCGCCGTGAAGTGAACGGGGTCA	14077
Oy	1277	TCGTCAAAGGGGTACTTTTCGTCGGGCGTTGCTCGATTAACCTTGGAAATGTCACAGATGGCTACG	13316
Dd	1408	ACGTCAAAGGGGTACTTTTCGTCGGGCGTTGCTCGATTAACCTTGGAAATGTCACAGATGGCTACG	14677
Oy	1337	GACCCAGATTGGCGGTCAAGCTTACACAGCTA	1367
Dd	1468	TGACGAGGTTGGGGGTTACGATATGGAATTA	1498
RESULT 6 AKI19546			
LOCUS	AKI19546		
DEFINITION	Oryza sativa (japonica cultivar-group)	CNA clone:001-207-A05, full insert sequence.	
ACCESSION	AKI19546		
VERSION	AKI19546.1	GI:37989169	
KEYWORDS	FLU-CDNA; oligo capping.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartidae; Oryzaceae; Oryza.		
AUTHORS	1		
REFERENCE	The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-Length cDNA Project Team;		
AUTHORS	Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Koijima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group.; Otono,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kuratski,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Naitaka,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Nikkura,D., Ikeda,R., Ishibiki,J., Kawamatsu,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN;		
AUTHORS	Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuoka,T., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Oono,N., Oga,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M., and Hayashizaki,Y.		
TITLE	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice		
JOURNAL	Science	301 (5631), 376-379	(2003)
MEDLINE	22752273		
PUBMED	12869764		
REFERENCE	2		
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hizamoto,K., Hirooka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Koijima,K., Koijima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kuwahara,C., Kuratski,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Morikami,K., Murata,M., Nagata,T., Nakahana,T., Nakamura,M., Namiki,T., Narikawa,R., Nikkura,J., Niishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,Y., Oono,N., Oga,Y., Otono,Y., Ryu,R., Saitoh,H., Sakai,C., Sasaki,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahita,S., Tanaka,T., Tomaru,A., Toyota,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S., and Yoshimura,A.		
TITLE	Collection, mapping, and annotation of 28K full-length cDNA clones from japonica rice		

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
3 (bases 1 to 1828)
Kikuchi,S.
Direct Submission
Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of
Agricultural Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail:skkuchienias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 32k full-length cDNA clones from japonica
rice.
URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
NIMS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K.,
Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
Ishikawa,M., Yamada,H., Ooka,H., Hotta,T., Kojima,K., Nemiki,T.,
Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T.,
Yamamoto,M. and Nakahama,Y.
FAIS Genome Sequencing & Analysis Group: Oono,Y., Iida,Y.,
Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M.,
Kobayashi,M., Kodama,T., Kuroaki,T., Kusumegi,T., Lu,M.,
Maeda,H., Miura,J., Mizuno,K., Nakikawa,R., Nikiura,J., Oka,M.,
Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsumoda,Y., Ueda,M.,
Xie,Q., Yokomizo,S., Yoshimura,A., Matubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
Hirooka,T., Hori,F., Iida,Y., Imamura,K., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
Kishikawa-Hirozane,C., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishii,K., Nomura,K., Numaaki,R., Ohno,M., Otsao,N.,
Ota,Y., Saitoh,H., Sakai,C., Sakai,K.,
Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K.,
Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M.,
Tagami-Takeda,Y., Tagawa,A., Takehashi,F.,
Takaku-Akihita,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
Yaeunishi,A. and Hayashizaki,Y.
Location/Qualifiers
1. 1828
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="001-207-A05"

ORIGIN

Query Match 19.1%; Score 278.2; DB 8; Length 1828;
Best Local Similarity 54.7%; Pred. No. 4.1e-42;
Matches 785; Conservative 0; Mismatches 608; Indels 42; Gaps 10

20 TGCCCAAGCACTTGAATGGGGCTTCGCAAGCGCGCTTACACAGATCGAAGGCGCGCTCA 79
Db TCCCGAGGGGTTCGCTTCGCGCACCGCTCCGCCGCTACACAGTACGAGGAGCTGTCA 267
80 AAGAAAGTGGCGCGCGCCGTCATCTTGGGACACGTAAGTCCACCTGAGCCATCGCGCA 139
Db AGGAGGAGCGGAGAGGCGCAGACCATCTGGGACACGTTGCCGACACCTTTGGAAAGATCA 327
140 CCAACGGGCGCAAGCGCATGTGCTTCGATCACTACCAACCGCTACGATAGGACCTTTG 199
Db CGGACTTGAAGAAAGCTGATGTTGCAATGATCACTACACCGTTTGAAGAGGATTAATAC 387
200 ATCTCTTACCAAGTACGCGCAAGCGCAAGCGCTTCTCCCTGTCGTGTCGCGCATCA 259
Db AGCTCATGCGACAGCATGGGAGATGATGCTATCGGTTCTTCATAGCATAGCTCAAAATCT 447
260 TTCCCTTCGGCGGCGAGCTGTGATCCCGTCAACGAGAGAGGAATGAGTTTACAGCAAC 319
Db ACCCAATATGATG-----TTGTGATCAATGATCAAGCTGTGATGACACATCAACAAGC 501
320 TGAATTGAGCCCTGTAGAGCGGGGTATCAGCCCTTGGGTGATCTTTGATCACTGGGATC 379

Db 502 TGATGATGCACTTTAGCAAAAGAAATTCAGCCATATGTGACTTACACTTACCACTGGAGC 561
Oy 380 TGGCTCAGGCGCTTCACGATGCTATGAGAGCTGGCTCAAGCTGGAAGAGTCCAGCTGG 439
Db 552 TCCCCAAGGCTTTAAAGCAAGAGGCTGGCTTGACAGGCA---GATAGTGAGC 618
Oy 440 ACTTTGACCGGTATGCGAGGTGTGCTTTGAACTTTTGGGGACCGAGTCCAGATGGA 499
Db 619 ATTTGCGGCGGTACGCGGAGAGTGTCTTCAAGGAGTTGGGGACAGGGTGAAGCACTGGA 678
Oy 500 TCACCATTAAGAGAGGCTTGAATTCAGGCTATCTATGATATGATCCCGGACGCAAGCCG 559
Db 679 TCACGCTTAAGAGCGGACAGAGGCTATCCAGGCTTAAGAGCGGCTTCCAGGCGCC 738
Oy 560 CGGCGAGAGGAGC-----ATTAAAGCACTCCACGAGGGGAAACACTGCACTGAGC 613
Db 739 CCGGCGGCTGTCTCCGTGTCTCCACTTCTATGAGAGGCGGCACTCCGCGACCGAGC 798
Oy 614 GTGGCTGTGTAAGAAAGGCGGAGATGATGAGCCATGCGCGCGGTGGCCGTCTTACAGCA 673
Db 799 CCTACGTGTGCGCCACCATTTCACTTCGCCCAAGCGCGCGCGCGGAGCATCTTACAGGA 858
Oy 674 GGGAGCTTGGCCCTCGGAAAGGCGGAGATCGGATCTGCTCAAGCGGCACTACTATG 733
Db 859 CAAGATACAGGCGACGAGAGCGGCGGCTTGGATAGCTTGAAGCTGATGTGTTCG 918
Oy 734 AGCCCTGGGACAGGATGAGCTCGGACAGAGAGCTGTGAGCGAGCGATGGAATTC 793
Db 919 AGCCGATGTCCAAACCCAGATC---GACATCGAGGCGGCAAGAGCGAGAGGTTTC 975
Oy 794 ACATTGGCTGTGTTGCCATCTCTTCTTGAAGAGAGATTCAGAGAGCATGAGA 853
Db 976 AGCTAGAGTGTGTTGCTGATCGTCTTCTT---CGGCACTACCGCGGAGCATGAGG 1032
Oy 854 AGCAGCTGGGAGAGGCTTCCAGCCCTCACTCCCGGAGATTGGCATCTCTCAATCCG 913
Db 1033 CGAGGCTGGGAGAGGCTGCGAGGTTCCAGGCGAGTGAAGCGCGCGCTGTCAAG--G 1089
Oy 914 GAGAGCCGACTTTCAGGAGTAATTAACAATCCAGTTGCGCGGCACTTAGAG 973
Db 1090 GGGCGCTGATTTGTCGCGATTAACCACTACACACTCTACTACAGAGGACAAACA 1149
Oy 974 GTCCCGTCCCGA---CGGACTATCTGGGCGCATCTCATAGCAACAGAGAT-- 1026
Db 1150 CCAACATCATCGGAGCATGTGCTAAACAACCTTGGAGACACCGGACCGTCAAGCTCC 1209
Oy 1027 ---AAGACGAGGCGCGCTTGGCGAGAGAGCGGCTCGCTGGTGGCGCTCGC 1081
Db 1210 CATTCAGAGAGAGGAGCAATTTGAGATAGGAGAAATTTGATATGCTGTACATTGTGC 1269
Oy 1082 CGGACATGTTCCGAGAGCATCTCGCCGCGGTGAAGCGCTGT--ACGGCAAGCCATCT 1138
Db 1270 CCCGAGGAGTGAAGAGCTGTGAATATGTCAAGGAAAGTACACAGCCCACTGT 1229
Oy 1139 ACATCACGAGAGAGGATGCCCGTGCCTGAGAGAGAGAAATGACGTGGAAGAGCCG 1198
Db 1330 ACATCATCTGAAAGAGGAGTGAATGACAGCAACACCGCTTCAATTCATCAAGAGACGCC 1389
Oy 1199 TCAAGACCCCTTCCGATCGSTACTTGTGACTGCACTTGAATTCGATTTCCAGGCGCA 1258
Db 1390 TCAAGAGCAGAGAGAGATCAAAATACACAAATGATCTCAACCAATCTGGCTTCCA 1449
Oy 1259 TTACCCAGAGAGCGGCTGTGTCAAGGGGTACTTTGCTGGGCGGTGGCTGATTAAGCTTG 1318
Db 1450 TCAAGAGAGAGCGGCTGTGAGTACGTGTGATCTTGGGTGTCTGTCTGAGCAACTGG 1509
Oy 1319 AATGCTCAGATGGCTACGAGCCAGATTCGCGCGTCAAGTTTCAAGACTA---CAACC 1375
Db 1510 AATGCTCAGATGGCTACCTCTGAGATTCGCGCTTCACTTGTGATTAAGAGATAACC 1569
Oy 1376 TCAAGCGCAGCGCCCAAGAGTGTGCTGTCTCAAGAGCATGTTTGGCGCGC 1430
Db 1570 TCAAGAGATACCCCAAGACTCGGTGAGTGTCAAGGCTCTCTGAAGACTTG 1624

RESULT 7
AY129294
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS
ORIGIN
Query Match 19.1%; Score 278.2; DB 8; Length 2032;
Best Local Similarity 54.7%; Pred. No. 4e-42;
Matches 785; Conservative 0; Mismatches 608; Indels 42; Gaps 10;
Oy 20 TGGCCACGACTTTGATGGGCTTGGCAAGCGCGCTTACAGATCGAAGCGCGCTGA 79
Db 290 TCCCCAGAGGAGTTCGCTTCGACCGCTCCGCGGACAGTACAGAGGAGCTGTGA 349
Oy 80 AAGAAAGTGGCGCGCGCGCTTCATCTGGGACAGCACTGACACCTGAGCGCATGGCGCA 139
Db 350 AGAGAGAGCGGAGAGGCGGACCACTTGGGACAGCTTGGCGCAACCTTTGAAAGATCA 409
Oy 140 CCAAGCGCGCAAGCGGAGTGTGCTTGGGATCACTACACCGCTACAGATGAGAGACTTTG 199
Db 410 CGGACTTCAAGCAATGCTGATGTTGAGTGTGATCACTACACCGTTTCAAGAGAGATAC 469
Oy 200 ATCTCTTGAACAGTACGCGGCAAGGCTTACCGCTTCTCTTGTGTGTGGCGATCA 259
Db 470 ACTCATGCGACAGATGGGAGTGTGATGCTGTATCGGTTCTGATACATGTGTCAAAATCT 529
Oy 260 TTCCTCGGCGGAGGCTGATCCGTCACAGAGAGGAGATGATTTTACAGCAAC 319
Db 530 ACCCAATGTG-----TTGCTCAAGTCAATCAAGCTGTGTGACCACTACAAAGC 583

QY 320 TGATTTGACGCGCTTTGAGCGGGGATTCACGCTTGGTGACTTTGATACCATCGGATC 379
 |||||
 Db 584 TGATCGATGCACTTCTAGCAAAAGAAATTGAGCAATATGAGCACTTACACATCGGGACC 643
 |||||
 QY 380 TGGCTCAGGCGCTTCACGATCGTATGAGAGCTGGCTCAACGTTGAAAGAGTCCAGCTGG 439
 |||||
 Db 644 TCCCCCAGGCGCTTGAAGACAAAGATCAAGGCTGGCTTGAACAGCA--GATAGTGACG 700
 |||||
 QY 440 ACTTTGAGCGGTATGCGAGGTGAGCTTTGAAAGCTTTGGGGACCGAGTCCAGAACTGGA 499
 |||||
 Db 701 ATTTCGCGGGTACGCGAAGCGCTTCAAGGAGTTCCGGGAGAGGAGTGAAGCACTGGA 760
 |||||
 QY 500 TCACCATCAACGAWCCCTGATTCAGGCACTTATGATATGACCCGCGAGCAACGCC 559
 |||||
 Db 761 TCAGCTCAACGAGCGGACACGCTGGCCATTCAGGGCTAGACGAGGGCTCCAGGCC 820
 |||||
 QY 560 CGGGCAGAGCAAGC-----ATTAAACAGCACTCCACCGAGGCAACCTGCACTGAGC 613
 |||||
 Db 821 CCGGCGCTGCTCTGCTGCTGCTCACTCACTCAAGGCGGCACTCCGCGCACCGAGC 880
 |||||
 QY 614 CGTGGCTGCTGGAAGGCCAGATTCATGAGCCATGCGCGCGCGTGGCGCTTACAGCA 673
 |||||
 Db 881 CTAACGTGCTGCGCCACCATCTTCTCGCCACGCGCGCGCGCAGCATTTACAGGA 940
 |||||
 QY 674 GGGACTTTGCGCCCTCGCAAAAGGCGCAGATCGGCACTCGCTCAACGCGCACTACTATG 733
 |||||
 Db 941 CAAATTCAGAGGCGACCGCAGACGCGGACCTTGGGATAGCTTGAAGTATGCTGCTTCG 1000
 |||||
 QY 734 AGCCCTTGGGACAGCAATGAGCTTCGGGACAGAGGCTGTCAGACGAGATGGAATTC 793
 |||||
 Db 1001 AGCCGATGTCCAAACACCAACGATC---GACATCGAGCGCGCGCAAGAGCGGAGGATTC 1057
 |||||
 QY 794 ACATGGCTGGTTTGGCAATCCCATCTTTGAAAGAGATATCCAGAGAGCATGAGA 853
 |||||
 Db 1058 AGCTAGATGATGTTGATGATCGTTCTTCTT---CGCGCATACCGCGCAGATGAGGG 1114
 |||||
 QY 854 AGCAGCTGGGCGAGAGGCTTCAGAGCCCTCACTCCCGCGGACTTGGCATCTCAATCCG 913
 |||||
 Db 1115 CGAGGCTGGGGAGAGGCTCCGAGGTTCAAGCGGAGTGAAGCGCGCTGCTCAAG---G 1171
 |||||
 QY 914 GAGAGACCGACTTCTACGCGATGAATTAACAATCCAGATTGCGCGCGCACTTAAGC 973
 |||||
 Db 1172 GGGCGCTGATGTTGCTGGGCTAAACACTACACACTACTACAGAGGCAACAAACA 1231
 |||||
 QY 974 GTCCCGTCCCGAGA-----CGGACTATCTCGGCGCATCATAGCAACGAGGAT-- 1026
 |||||
 Db 1232 CCAACATCATCGGACATGCTCAACAAACCTTGGAGACACCGGACCGCTGAGCTCC 1291
 |||||
 QY 1027 -----AAGGACGCGAGCGCGGTTGGGAGAGAGCGGCTCGCTGGCTGCGCTCGCC 1081
 |||||
 Db 1292 CATTCAAGAAATGGGAGCCAAATTTGGAATAGGCGAAATTCGATATGGCTGTAATGTGC 1351
 |||||
 QY 1082 CGGACATGTTCCGAGACATCTCGCCGCGGTGAAGCGCTGT--ACGCGAAGCCCATCT 1138
 |||||
 Db 1352 CCGGAGGATGAGGAGCTGTAATGTAATGTAAGGTAACAAGCCCAACAGTGT 1411
 |||||
 QY 1139 ACATCAGCGAAGAGATGCGCGGCTGCGAGAGGAAACATACGTCGAGAGAGCGCG 1198
 |||||
 Db 1412 ACATCAGTGAAGAAAGGAGTGAAGACAGCAACCCGTTCAATTCATCAAGAGACGCC 1471
 |||||
 QY 1199 TCACGAGACCCCTTCGCGATCGGTAATTTGACTGCACTTGAATTCGATTTCCAAAGCCA 1258
 |||||
 Db 1472 TCAGAGGACAGCAAGAGATCAAAATCAACAATGACTACCTCAACAACTGGCTGCTTCCA 1511
 |||||
 QY 1259 TTACCCAGAGACGCGCTGCTCAAGGGGTAATTTGCTGGGCGCTTCTCGATTAATTTG 1318
 |||||
 Db 1532 TCAGAGGAGACGCGGTGCGAGTACGTGGGTACTTCGCGGTGCTCTGCTGGAACAATGGG 1591
 |||||
 QY 1319 AATGCTCAGATGCTACGAGACCAAGATTCGCGCTCAAGTTTCAAGACTAC---CAACCC 1375
 |||||
 Db 1592 AGTGGGCGCGGATCTCTCGAGATTCGGGCTCTACTTCTGTAAGACTCAAGAGATTAAC 1651
 |||||
 QY 1376 TCAAGCCACGCGCAAGAGATCTGCTGCTCAAGAGCATGTTTGGCGCGCG 1430
 |||||

Db 1652 TCAGAGATACCCCAAGAACTCGCTGAGTGTTCAGAGCGCTCTGAAGACCTG 1706
 |||||
 RESULT 8
 AF268911
 LOCUS
 DEFINITION
 Aspergillus niger beta-glucosidase precursor (bg11) gene, complete cds.
 ACCESSION
 AF268911
 VERSION
 AF268911.2
 KEYWORDS
 GI:8572535
 ORGANISM
 Aspergillus niger
 Aspergillus niger
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 REFERENCE
 1 (bases 1 to 1988)
 A1,Y.-C. and Meng,F.-M.
 Molecular cloning and sequencing of a beta-glucosidase gene from
 Aspergillus niger AMS11
 Unpublished
 JOURNAL
 2 (bases 1 to 1988)
 A1,Y.-C. and Meng,F.-M.
 Direct Submission
 Submitted (17-MAY-2000) Department of Biochemistry, Zhongshan
 University, 135 Xingang Road, Guangzhou, Guangdong 510275, P. R.
 China
 3 (bases 1 to 1988)
 A1,Y.-C. and Meng,F.-M.
 Direct Submission
 Submitted (19-JUN-2000) Department of Biochemistry, Zhongshan
 University, 135 Xingang Road, Guangzhou, Guangdong 510275, P. R.
 China
 REMARK
 COMMENT
 On Jun 19, 2000 this sequence version replaced gi:8248474.
 FEATURES
 source
 1..1988
 /organism="Aspergillus niger"
 /mol_type="genomic DNA"
 /strain="AMS11"
 /db_xref="taxon:5061"
 <496..>1858
 /gene="bg11"
 join(<496..536,595..>1858)
 /gene="bg11"
 /product="beta-glucosidase precursor"
 join(496..536,595..1858)
 /gene="bg11"
 /codon_start=1
 /product="beta-glucosidase precursor"
 /protein_id="AA074209.2"
 /db_xref="GI:8572535"
 /translation="MLPKDQMGFAKAAVQIEGAVDQDGRSPISWDTFCAOPGKIADG
 SSGVTACDSYNTAEADIALKSLGAKSYFSSRIPEGGRDAVAVQAGIDHYVKFVD
 DLDADGITPTLTFLHMDLHORYGALNRPFLDEPNYRVVRFALPKRVNNEPLC
 SAIRGVGSGFAPRGROSTSPMTVGHNIIVAHGRAVAKAYADDFKPSAGDOIGIVNG
 DPTVPMADADPAKRLERLPTFAFDPIYIGDVPASMRQIDGRLPPTPEERAAVAG
 SDFYGMNHTSYIRHRSPPASADPTGVNDVLFTRKQNCICGPEPTQSPMLRPCAG
 RFDPLWTSRYSPEPIVYENGTSLKGSBDLENEKILEDFFVKYNYETIRAMVAV
 ELDGAVRFGVTVVDYENGQKRPKKSASLKLPLDELIAAA"
 join(496..536,595)
 /gene="bg11"
 /gene="bg11"
 596..1855
 /gene="bg11"
 /product="beta-glucosidase"
 ORIGIN
 Query Match 17.4%; Score 253.2; DB 8; Length 1988;
 Best Local Similarity 56.0%; Pred. No. 1.9e-37;
 Matches 708; Conservative 1; Mismatches 474; Indels 81; Gaps 9;
 QY 45 CGAAGCGCGCGCTACAGATCGAAGCGCGCTCAAGAAAGAGTGGCGCGCGCTTCAT 104
 |||||

Db 583 CCCGTCGACAGCTACCAAGATGAGGGCCGCTGCAGACAGAGCGGCGCGGCCAGCAT 642
 Oy 105 CTGGAGCAGCTACTGCGCAGCTGAGCCCATTCGCGACCAACGGCGCCCAACGGCGATGTGGC 164
 Db 643 CTGGGACAGCTTCTGCGCGGAGCCCGGGAAGATTCGCCACGCGCTGTGGGCGTGAACGGC 702
 Oy 165 TTGGCATCACTACACCGCGTACAGATGAGGACTTTGATCTTTGACCAAGTACGGCGCAA 224
 Db 703 GTGCACTCTGTACAAACCGCACCGGAGAGACTTGGCTCAAGTCCGCTCGGGGGCCAA 762
 Oy 225 GAGCCATCCGCTTCTGTCGTGGTGGGGATCATTCCTCCGCGCGGAGGCTGGATCC 284
 Db 763 GAGCTACCGCTTCTTCATCTGCT-----CGCGCATCCCGAGGGCGGCGCGCGGATGC 816
 Oy 285 CGTCAACGAGGAGGGAATTGATTTTACAGCAAACTGATTAAGCGCCCTTTGAGGCGGG 344
 Db 817 CGTCAACGAGGCGGCGCATCGAACACTAGTCAAGTTCGTGACAGCATCTGCTGACGCGG 876
 Oy 345 TATCAGCGCTTGGGCTGATCTTTGATCACTGGGATTCGCTCAGGCGCTTCAAGATCGCTA 404
 Db 877 CATCAGCGCTTTCATCACCCCTTCTTCACCTGGGACCT-----GCTGCATCAGCGGTA 927
 Oy 405 TGAAGCTGCTCAACGTGAAAGAGGTCCAGCTGAGCTTTGAGCGGTATGAGAGTTGTG 464
 Db 928 CGGGGGGCTGCTGAACCGCACCGAGTTCGCTGACCTTGAATACTAGCGCCGCTCAT 987
 Oy 465 CTTGGAACGTTTGGGAGCAGAGTCCAGACTGATCACTCAACGAMCCCTGATTTCA 524
 Db 988 GTT-----CAGGGCGCTGCGCAAGGTGCGCAACTGGAACGAGCGCGTGTCTC 1035
 Oy 525 GGGCATCTATGATATGATCCACCGGAGCAAGCCCGGGCAGAGCAGCATTAACAGCA 584
 Db 1036 GGGCATCTCCGGGCTACGGCTCCGGAGTGTGCCCCCGCGCGAGAGCA----- 1085
 Oy 585 CTCACCGAGGGCAACACTGCGACCTGCGCTGCTGAAAGAGCCGAGATCATGAG 644
 Db 1086 -----CCTCGAGCCGTGACGCTCGGCGCAACAATCTCTGTGC 1125
 Oy 645 CCATGCCCGCGCCGTGCGCTTACAGAGGAGACTT---TGCCCTTCGCAAAAGGGCA 701
 Db 1126 CCAGCGCGCGCCGCTCAAGGCGTACCGGAGCATTTCAAGCCCGCAGCGCGAGCGGCA 1185
 Oy 702 GATGGCATCTTCGCTCAACCGGCGACTACTAGAGCCCTGGGACAGCAATGAGCTCGGGA 761
 Db 1186 GATGGCATCTTCGCTCAACCGGCGACTACTAGAGCCCTGGGAGCGCGCGAGCGGCGGA 1245
 Oy 762 CAAGAGGCTGCTGAGCAGCGAGATGATTTCACTTGGCTGTGTTTGGCAATCCCATCTT 821
 Db 1246 CAAGAG-----CGCTCGAGTCTTTCACGCGCTGTGTTCCGGATCCCATCTA 1293
 Oy 822 CTGGAAGAGGACTATTCAGAGAGCATGAGAGCAAGCAAGCTGGGCGAGAGGCTTCCAGCCCT 881
 Db 1294 CTTGGGAC---GACTTACCGCGCGTGCATCGCAAGAGCTGGGCGACCGGCTGCGGACCTT 1350
 Oy 882 CACTCCCGCGGACTTTGGCATCTCTCAATGCCGAGAGACCGACTTTTACCGCATGAATTA 941
 Db 1351 TAGCGCGAGGAGCGCGCTCTGTCA---GCGCTCCAAACGACTTTTACCGCATGAACCA 1407
 Oy 942 CTACACATCCAGTTCGCGCGCCACTTACGAGTCCGCTCCCGGAGAGCGACTATCTTCG 1001
 Db 1408 CTACACGCTCAACTACATCCGCGACCGAGAGTCCGCGCTCCCGCGACGACACCGCTCG 1467
 Oy 1002 CGGCATTCATGAGCAGCAGGAATAAGAGCGGAGCGGCGCTTGGCGAGAGAGCGGCT 1061
 Db 1468 CAAGCTGACGCTGCTTTCACCAACAGAGGAGCAATGACATCGAGCCCGCGAGAGCGCATC 1527
 Oy 1062 CGCGTGGCTGCGCTTCGCGCGGACATTTCCGAGAGCATTCGCGCGGGGTATCGGCT 1121
 Db 1528 CCGCTGGCTGGCGCCCTGTGCGCGGCTTCGCGACTTCTGTGTGTGAGCCAGCAAGAG 1587
 Oy 1122 GTACGG---CAAGCCCATCTACATACAGAGAGCGATGCGGCTGCGCTTGGAGGAGAA 1178
 Db 1588 GTACGGCTCCCGCCCATCTACGTGACGAGAACCGCACAGCATCAAGGGCGAGAGCGA 1647

Oy 1179 CATGACGTGCGAGAGGCGCTGCAACGACCCCTTCCGATCCGTACTTGTGATCGCACTT 1238
 Db 1648 CTTGCCAAGAGAGATTTCTCGAAGATGACTTCAAGGCTCAAGTACTATTAACGATACAT 1707
 Oy 1239 GGACTGATTTCCAGGCCATTTACCGAGAGCGCGCTGTGTGTAAGGGGATTTGGCTG 1298
 Db 1708 CCGTGCATGTTTACCGCGCTGAGAGCTGACCGGGGTTCACGTGAGTGGGGGTTACGTA 1767
 Oy 1299 GCGG 1302
 Db 1768 TGTG 1771
 RESULT 9
 AK067001
 LOCUS
 DEFINITION
 Oryza sativa (japonica cultivar-group) cDNA clone:J013092F09, full insert sequence.
 ACCESSION
 AK067001
 VERSION
 GI:32977019
 KEYWORDS
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE
 AUTHORS
 The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Ohtomo, Y., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kuroseki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN; Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishi, Y., Itoh, M., Kagawa, I., Kondo, S., Komono, H., Miyazaki, A., Otsu, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.
 Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
 Science 301 (5631), 376-379 (2003)
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 2 (bases 1 to 2045)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizaki, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, P., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imanura, K., Imotani, K., Ishibiki, J., Ishi, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Komono, H., Kouda, M., Koya, S., Kurihara, C., Kuroseki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Otsu, N., Ota, Y., Ohtomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takehashi, F., Takaku-Akashita, S., Tanaka, T., Tomaru, A., Taya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.
 Direct Submission

TITLE

JOURNAL

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kennondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT

This clone is one of the 28k full-length cDNA clones from japonica rice.

URL: http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project_Team_Kikuchi_S_Sato_K_Negata_T_Kawagashira_N_Doi_K_Kishimoto_N_Yazaki_J_Ishikawa_M_Yamada_H_Ooka_H_Hotta_I_Kojima_K_Namiki_T_Ohmeda_E_Yahagi_W_Suzuki_K_Li_C_Ohtsuki_K_Shishiki_T_Yamanoto_M

FAIS Genome Sequencing & Analysis Group: Otsuno, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurobaki, T., Kusumegi, T., Lu, M., Masuda, H., Mura, J., Mizuno, K., Naitaka, R., Naitaka, J., Oka, M., Ryu, Q., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numasaki, R., Ono, M., Otsu, N., Oda, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shihara, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takata-Kahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.

FEATURES

Source

1. 2045
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultiivar="Nipponbare"
/db_xref="taxon:3947"
/clone="J013092F09"

ORIGIN

Query Match 17.1%; Score 248.6; DB 8; Length 2045;
Best Local Similarity 52.3%; Pred. No. 1.4e-36;
Matches 730; Conservative 1; Mismatches 632; Indels 33; Gaps 7;

20 TGGCCACGACTTGAATGGGGCTTGGCAAGCGCCGCTACCAAGTGAAGCGCGCTCA 79
Db 256 TCCCGAGGACTTCATCTCGGCAACGGCTCCGGGCTTACAGTACGAGGCGCTTGA 315
Qy 80 AAGAAGTGGCGCGCGCGCTCCATCTGGGACAGCTACCTGACCTGGAGCCATCGGCA 139
Db 316 ACAGAGCGCGCAAGGCGCTGAGCATCTGGGACAGCTACCTGACCTGAGGCGCAATTG 375
Qy 140 CCAAGCGCGCAAGGCGCTGAGCATCTGGGACAGCTACCTGAGGCGCTTGG 199
Db 376 AAGATGTAGCAACGGGAGTGTAGCACTTTTACATCGCTACAGAGAGATTTGA 435
Qy 200 ATCTCTTGACCAATGACGGCGCAAGGCTTACCGCTTCTCTTGTCTGTGCGGATCA 259
Db 436 ACTTCGTGACGACATGACATGAGAGCGCTTCGCTTCTCATTTGTGAGAGAGATTC 495
Qy 260 TTGCGCGCGCGGAGGCTGATCCGCTCAACGAGGAGGATTTGATTTACAGCAAC 319
Db 496 TGGCAATGAGACATCATGAGGAGATCAACAAAGGGGATTTGCTTTTACAAACGCC 555
Qy 320 TGAATGACGCGCTTGAAGCGGGGTATCAAGCGCTTGGTGAATTTTACATCGGATC 379
Db 556 TGAATGACGAGTATCTTGAAGGTTGAAGCATTTGTCAATTTTCAATTTTGACA 615
Qy 380 TGCCTCAGGCGCTTACGATCGCTATGAGAGCTGGCTCAAGTGAAGAGCTCCAGCTG 439
Db 616 CCCACAGGCTGTGAGAGCAAAATACCGAGGCTTCTCA---GTGAATAATTTGAAAG 672

Qy 440 ACTTGAAGCGGTATGCGAGTTGCTTTGAACGTTTGGGAGCCGATTCAGAACTGGA 499
Db 673 ATTTCGTGACATATGCGAGCTGTGCTTCCGTAGTTCCGCGACCGGTTAAATCATGGA 732
Qy 500 TCACCATCAACGAGCCCTGATTCAGCCATCTATGATATGCTACCCGAGCAAGCCCC 559
Db 733 ACAGTTTCAACGAGCCGATGATTTCTTGGCGCGCGGCTTACGCGCTCGGACCAAGCCCC 792
Qy 560 CCGGAGGAGC-----AGATTAAAGAGCTCCACGAGGCAACATGCCACTGAGC 613
Db 793 CCGGCGCTGCTGCGGTAGCTTCCAGAAATGCGCCCCCGGAGCTCGGAGAACGAGC 852
Qy 614 CTTGGCTGCTGGAAGAGCCAGATCATGAGCCATCGCGCGGCTGAGCTTACAGCA 673
Db 853 CTTAGCTGCGCGGCAACACTGTCTGCGCCACCGGAGGCGCTTACCGCC 912
Qy 674 GGAAGCTTTCGCGCGTGGCAAGAGGCGCAATGCGCATCTGCTCAACGCGCACTATG 733
Db 913 AAAAGTACAGCGAGCGAGAGAGAGGAGGAGATCGGCATCAAGGATGTCGACTGTTG 972
Qy 734 AGCCCTGGAGACGATGAGCTTCGGGACAAAGAGCTGCTGAGCGAGCATGAAATTC 793
Db 973 TGGCTT---ACAGCGAGCGCGCGCGGACAGACGCGCTGAGGCGAGCTTGACTTCA 1029
Qy 794 ACATTCGCTGCTTGGCAATCCCATCTTCTTGAAGAGAGCTATCCAGAGAGATGAGA 853
Db 1030 TGTACGATGCTTATGACAGCCGATGTGTT---CGGCGATCAACCGGAGCATGAGA 1086
Qy 854 AGCAGCTGGCGAGAGGCTTTCAGCCCTCACTCCGCGGACTTTCATCTCAATGCGC 913
Db 1087 AGTTGGCTGAGCGGCTGCGGCTGCGCAAAATTCACAGCGAGTGGAGTTGTGA---GG 1143
Qy 914 GAGAGACGACTTCTACGCGATTAATTAATCAATCCAGTTCCG-----C 961
Db 1144 GATCTACGACTTCTACGCGCTCACTACCTACCAACCACTACGCTTACAGGCTTAC 1203
Qy 962 GCCACTTACGAGGCTCCGCTCCCGGAGAGAGAGCTATCTCGGCGCATTCAGAGCAGC 1021
Db 1204 GAGCGCGAGCAAGCTTCAACCGGCTTACCTGATCACTAGCACTAGGATTAATCAAC 1263
Qy 1022 AGAATTAAGACGAGCAGCCCGCTTGGGAGAGAGAGCGGCTTGGCTGCGC 1081
Db 1264 ACAGCAAGCGGCTCCGATGCGTCCGCGGCTTTACAAAGATCTTCACTACGCGC 1323
Qy 1082 CGGACATGTTCCGAAAGCATCTGCGCGGCTGAGCGCTTACGCGAGCCATCTACA 1141
Db 1324 CGGCTCTGCGTGAAGCTGCTTGTATACCAAGAGAAATGACATGACATCTACA 1383
Qy 1142 TCACGAGAAAGGATGCGCGTGGCTGAGAGAGAAATGAGTGGAGAGAGCGCTCA 1201
Db 1384 TTGCAAGAAAGGAGCAGCAGAGAGAGCAACAGCAGATCTTATGCGGAGCGCTCA 1443
Qy 1202 ACAGACCTTTCGCGATCCGCTTACTTGAATCTGACCTTGAATTTCCAGGCGCATTA 1261
Db 1444 AGGATGACAAAGCAATATCTTCCATACAGATCTCGCT---TCACTCAGCTGCA 1500
Qy 1262 CCAAGAGCGGCTGCTGCTCAAGGATCTTTGCGTGGGCTTCTGATTAATTTGAAT 1321
Db 1501 TCAAGAGAGGAGTGAAGGCTTACTTCACTTCACTGAGAGCTTCACTGAGT 1560
Qy 1322 GGTCAAGTGGCTTACGAGACCCAGATGGGCTGACGTTCAAGACTACACACCTCAAGC 1381
Db 1561 GGGGCGAGCGCTTACCGGCGCTTGGGCTTCACTTACGTCGACCGGAAATCTTCAAG 1620
Qy 1382 GCAAGCCCAAGAGTC 1397
Db 1621 GTATGCAAGAGTC 1636

RESULT 10
AF439322
LOCUS AF439322 2068 bp DNA linear PLN 20-JUN-2002

DEFINITION Talaromyces emersonii beta-glucosidase 1 (bg1) gene, complete cds.
ACCESSION AF439322 GI:21493049
VERSION AF439322.3
KEYWORDS
SOURCE Talaromyces emersonii
ORGANISM Talaromyces emersonii
Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Talaromyces.
REFERENCE 1 (bases 1 to 2068)
AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.
TITLE Molecular cloning and expression analysis of beta glucosidase from Talaromyces emersonii
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2068)
AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2001) Biochemistry, National University of Ireland, Galway, Ireland
REFERENCE 3 (bases 1 to 2068)
AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2002) Biochemistry, National University of Ireland, Galway, Ireland
REMARK Sequence update by submitter
AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Biochemistry, National University of Ireland, Galway, Ireland
REMARK Sequence update by submitter
COMMENT On Jun 20, 2002 this sequence version replaced gi:21284362.
FEATURES
source
location/Qualifiers
1..2068
/organism="Talaromyces emersonii"
/mol_type="genomic DNA"
/db_xref="taxon:68825"
<304..>2068
/gene="bg1"
/join(<304..383,472..663,732..1204,1258..1843,1930..>2068)
/gene="bg1"
/product="beta-glucosidase 1"
/join(304..383,472..663,732..1204,1258..1843,1930..2068)
/gene="bg1"
/EC_number="3.2.1.21"
/note="glycosyl hydrolase"
/codon_start=1
/product="beta-glucosidase 1"
/protein_id="AAL34084.2"
/db_xref="GI:21284363"
/translation="MTTETPTTTTSTLPPDFLWGFATASVQIEGAFEDDGLPSIM
DPSKTPGVKVDGNGDVACDSYHRTGEDIILKKYGAKIYRFSLSWRIIPGGRND
PINEKGLPFYSKFLDDLAAGIEPVYITFHNDLPDELMKRGIMANKKEPAPADANYA
RVFENALGSKVHMTTENEPCSSVYLGNTGKHAERGTSDRKTSPEGGTBEPMIVGH
NLTVAHGTVVDIYRREFEKOGEIGITLNDMAEPWPENPADVEACDRKIEPAISW
FADPIYHGKTPDSVWQGLDRLEKPTPEIAIFVHSGNSFYGMNHCENYINRKGEPD
PEDIAGNLIIMEDKNGNIPGETCEMLPEPLGFRLLKMLADRYNNPKIYIYENG
TSVGESEDKPLEEVNDEFERYQYVDYGAADVADAOGVAVKAVMAISLIDNPEWSE
GRRSRFGVTVYDYNKGQRIKKSALVIGELFNKTRKE"

ORIGIN
Query Match 17.0%; Score 247.8; DB 8; Length 2068;
Best Local Similarity 53.2%; Pred. No. 2e-36;
Matches 738; Conservative 1; Mismatches 518; Indels 130; Gaps 5;

QY 57 CTACCAAGATCGAAGCGCCGTCAGAAAGAGTGGCGCGCCGTCATCTGGGAACGTA 116
DB 472 CTACCAAGATCGAAGCGCGCTTCGACAGAGATGCGCTGCTCGATATGGGAACATT 531
QY 117 CTGCAACTGAGCGCATGCGGACCAAGCGCGCAACGCGCATGGCTTGGATCACTA 176
DB 532 CAGCAAGACGCCCGGCAAGGTGAGAGACCAATGGCGATGTGCGCTGCACTCGTA 591
QY 177 CCACGCTACGATGAGACTTGTATCTTGAACAAGTACGCGGCAAGGCGCTACCGCTT 236

DB 592 CCACCGTACCGGTGAGAGACATTGATATTTGAAAGAAATCGGTGCCAAGATCTATCGCTT 651
QY 237 CTCCTTGTCTGTGTCG-----252
DB 652 CTCGTGTCTCGGTGTGTCTGTATGTCTGCTCTTCTTATTTTGTCTGTCTGTTTTTT 711
QY 253 -----CGGATCATTTCCCTCGCGGCAAGGCTGATCCCGTC 288
DB 712 TGTCTGACAGAACATTCACAGCCACGATATCCCTCTCGGCGGGCAAGACCCCATC 771
QY 289 AACGAGAGGGAATTGAGTTTTCACGAAACTGATTCACCCCTGTTGAGCGGGGTATC 348
DB 772 AATGAAAGGCGCTGACGTTCTACTCAAGTCTTGATATCTTACAGCGCGCGGATC 831
QY 349 ACGCTTGGGTGACTTTGTACCACTGGGATCTGCTAGCGGCTTACGATTCGATGGA 408
DB 832 GAGCCCTTGTCAAGCTCTTCCACTGGGATCTGCGACGAGCTGATGAAGCGGTACGG 891
QY 409 GGCTGCTCAACGTGNAAGAGTCCAGCTGAGCTTTGAGCGGTATGCGAGGTGTGCTT 468
DB 892 GGCATGCTCAAGAGAGAGTGTGTGGCGGACTAGCCAACTATATGCCGTGTCTTC 951
QY 469 GAACGTTTGGGACCGAGTCCAGAACTGATCACTCAACGAMCCCTGATTCAGGCC 528
DB 952 AACGCGCTCGGTGAAAGTCAAGCACTGATCACTTCAACGAGCCCTGATGATTC 1011
QY 529 ATCTATGATATGCCACCGGACAGAACGCCCGGGGAGAGACATTAACAAGCACTCC 588
DB 1012 GTGCTGCGCAACAACCGGACAGACGCCCGGGGAGAGACATTCGAGGAAGT 1071
QY 589 ACCGAGGGAACCTGCACTGAGCCGTGCTGTGTAAGGCGCCAGATCATAGGCAT 648
DB 1072 CCCGAGGGAAGCGGCACTCGGAGACCTTGGATGTGTGTACAACTTTCGTGTCTAT 1131
QY 649 GCCCGCGCTGCGCTGTACAGAGGACTTTCGCCCTCGCAAAAGGCGCATATGCGC 708
DB 1132 GGTACCGTTGTGTGACATCTATCGAGAGAGTTCAGAAAGAACAGGCGGGGAATCGGA 1191
QY 709 ATCTGCTCAAC-----720
DB 1192 ATCAACCTCAACGGTAGGCTCGCGCTCATCTCGATTCGACTTATGTCTTACCGCATCT 1251
QY 721 -----GGCGACTACTATAGAGCCCTGAGACAGCAATGAGCCTCGGAGCAAGAGGCTCTG 775
DB 1252 ATCTAAGGAGATGCGCGGAACATGAGTCCGAAACCAAGCAGACGTGCAACCTGCG 1311
QY 776 AGCGACGATGGAATTTTCACATTGCTGTGTGCAATCCCATTTCTTGAAGAAGACT 835
DB 1312 ACCGAAAGATCGAATTCGCCATCTCGTGTTCGGGACCCGATCTAC--CACGGGAAT 1368
QY 836 ATTCAGAGAGCATGAAGAGAGCTGGGCGGAGAGGCTTCCAGCCTCACTCCCGGACT 895
DB 1369 ACCGGAAGAGATGTGAAACAGTTGGGACCGCTGTCCGAAATTCAACCGCGGAAGAAA 1428
QY 896 TTGCGATCTTCAATGCGGAGAGACGACTTCTACGCGATGAATTACTACATCCAGT 955
DB 1429 TCGCTTGTGTCAAGGAGC--AACGATTTTACGGGATGAACACTACGCAAAACT 1485
QY 956 TCGCGCGCACTAGAGCGTCCGTCGCCGAGAGAGACTATCTGGCGCCATTCATAGC 1015
DB 1486 ACATTCGGAACCGGACCGGCAACAGACCGGAGAGCAATTGTGCAACTGTGACATCC 1545
QY 1016 ACCAGAGAAATTAAGAGCGGACCGCGTGTGCGAGAGAGCGGCTCGCTGCGTGCCT 1075
DB 1546 TGAATGAGAGACAAAGACCGCAACCCCATCGGCCCCGAGACCAATGCAAGTGTGCGGC 1605
QY 1076 CTTGCGCGGACATGTTCCGAAGCATCTGCGCGGAGTACAGGCTGTAGCGCAAGCC- 1134
DB 1606 CGTTCCCGCTGGGGTTGAGAACTGTGAAGTGTGACAGACCGGTACAAACCCCA 1665
QY 1135 --ATCTACATACCGGAGAGCGAGTCCGTCCTGTGAGAGAGAAACATGACGTGCGAGG 1192

Db 1666 AGATCTACGTGACCGAGAACCGGACGCTCAAGGCGGAGAGCCGCTCGAGG 1725
Oy 1193 AGGCGCTCAACGACCCCTTCGCGATCCGTAATTGACTGCACTTGAGCTGATTTCA 1252
Db 1726 AGGTGCTCAACGAGAGTTCGCGCTGCACTACCGGAGCTATATGCGCGATGCTGG 1785
Oy 1253 AGGCGATTAACGACGCGGCTGCTGCTCAAGGGGATCTTTCGCGGCGTTCGATA 1312
Db 1786 ATGCTGTGACGACGAGATGGGGTCAAGCTTCAATGCTTGAAGTTTGTATAGT 1845
Oy 1313 ACTTGA 1319
Db 1846 AAGTGA 1852

RESULT 11
LOCUS AK654334 1404 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 4204 from Patent WO03000898.
ACCESSION AK654334
VERSION AK654334.1 GI:29157148
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1
AUTHORS Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
Kategrist,I.F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 4204 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
SOURCE 1..1404
Location/Qualifiers
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 16.9%; Score 246; DB 6; Length 1404;
Best Local Similarity 54.5%; Pred. No. 4,6e-36;
Matches 725; Conservative 0; Mismatches 565; Indels 40; Gaps 10;

Db 497 ACAGGGTGAAGCACTGGATATCAGCTCAAGAGCCGACACGATGGCCATTCAGGGCTACG 556
Oy 542 CCAACCGGACCAACGCGCCCGGAGAGAGCATTA-----ACAAGCATCCACGAGG 595
Db 557 ACCGAGGGCTCCAGGCGCCCGCGCTGCTCGTGTCTCTCCACTCTATATGAAAGCGG 616
Oy 596 GCAACACTGCACTGAGCGGTGCTGTGAAAGGCCAGATCATGAGCATGCCGCG 655
Db 617 GCAATCTCGGACCGAGCCCTAGCTGTGCGCCACCATTTCACTCTCGCCACGCGCGG 676
Oy 656 CGGTGCGCTTACAGACGAGGACTTTGCGCCCTTCGCAAAAAGGCGAGTGGCATCTCG 715
Db 677 CCGCCAGCATCTACAGGACAAATATCAAGGCGACGCAAGCGGCGAGCTGGATAGCT 726
Oy 716 TCAACGGCACTACTATGAGCCCTGGGACAGCAATAGCCTCGGACAAAGAGCTGCTG 775
Db 737 TCGACGTGATGTGTTCGAGCCGATGTCCAAACACAGAT--CGACATCGAGGCGGCA 793
Oy 776 AGGAGCGATGAATTTTCACTTGGCTGTTTCCATCCCATCTTCTTGAAGAGACT 835
Db 794 AGAGAGCGCAGAGCTTTCAGCTAGATGTTCCTGATCCGTTCTT---CGGGAGCT 850
Oy 836 ATCCAGAGACATGAAGAGACAGCTGCGGAGAGGCTTCAGGCTTCACTCCGCGACT 895
Db 851 ACCCGGACATGAGAGGCGAGGCTGGGGAGAGGCTGCCAGGTTCAAGCGGATGAGG 910
Oy 896 TTGCGATCTCATATGCCGAGAGACCGCATTTCTACGGCATGAATTCTACATCCCACT 955
Db 911 CCGCCGTCTGCAAG---GGGGCGCTGGATTTCTGTGGCATAAACAATACACCACTACT 967
Oy 956 TCGCGGCGCACTAGACGAGTCCGTCGCCGAGAG-----CGACTATCTCGGCGCCATCCA 1010
Db 968 ACACGAGGACAAACAAACCAACATATCGGACATTTGCTCAACACACCTTGGACACA 1027
Oy 1011 TGACACCAAGAGAAAT-----AAGACGCGACGCCCTTGGCGAGAGAGCGGCTCG 1063
Db 1028 CCGGACCGTCACTCCCATTCATGAATGGAGAACCAATTGAGATAGGGCAAAATTGCA 1087
Oy 1064 CTGCGCTGCGCTCTGCGCGGACATGTTCCGGAAGCATCTCGCCGAGTATAGGCTGT 1123
Db 1088 TATGCGTGTATCATTTGCCCCGAGGATGAGAGCTGTATGATATGTCMAAGAAAGT 1147
Oy 1124 ---ACGCAAGCCCATCTACATCACCAGAGAACGATGCCGCTCTGAGAGAGACA 1180
Db 1148 ACAACAGCCCAACAGCTATCATCTGAACCAAGGATGATGACAGACACACCTTTCA 1207
Oy 1181 TGACGTGCGAGAGGCGCTCAAGACCCCTTCGCAATCCGATCTTGAATCGCACTTGG 1240
Db 1208 TTTCATCAAGAGACGCGCTCAAGAGACAGAAAGATCAAAATGACATGACTTCA 1267
Oy 1241 ACTCGATTTCAAGGCGATTAACCAAGAGCGGCTGCTGTCAAGGGGACTTTGCGTGG 1300
Db 1268 CCAATCTGCTGCTTTCATTCMAAGAGACGCGGTGCACATGATCGTGGGTCTTGG 1327
Oy 1301 CGTTGCTGATTAATTGAA--TGGTCAATGCTTACGAGACCCGATTTGCGGCTGCTG 1359
Db 1328 CTCTGCTGCAACTGGGAATTGGCGGCGGATCTCTCGAGATTCCGGGCTCTACTTC 1387
Oy 1360 ACAGACTTACA 1369
Db 1388 GTGAGACTACA 1397

RESULT 12
AK067231 2148 bp mRNA linear PLN 24-JUL-2003
LOCUS AK067231
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013095012, full
insert sequence.
ACCESSION AK067231
VERSION AK067231.1 GI:32977249
KEYWORDS FLI cDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (Japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharidoideae; Oryzaceae; Oryza.

REFERENCE AUTHORS

the Rice Full-length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-length cDNA Project Team:
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H.,
Hotta, I., Kohjima, K., Namiki, T., Ohneda, E., Yehagi, W., Suzuki, K., Iji, C.,
Ohtsuka, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group: Otonari, Y., Murakami, K.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Naitakawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Mikiura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Nakura, J.,
Kusumegi, T., Oka, M., Ryuu, R., Ueda, M., Matsubara, K., RIKEN:
Kawaji, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imocani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Komu, H., Miyazaki, A., Oosato, N., Oka, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
Japanese rice
Science 301 (5631), 376-379 (2003)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

2 (bases 1 to 2148)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanasaka, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, T., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, S., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Komio, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Matsuda, H., Matsubara, K., Matsumura, T., Mura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Nariikawa, R., Nihura, J., Nishi, K., Nomura, K., Nunasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Oosato, N., Oda, Y., Otomo, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Saito, H., Saeki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akita, S., Tanaka, T., Tomaru, A., Toyota, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, kikuchi@nias.affrc.go.jp)

COMMENT

URL: <http://cdna01.dna.affrc.go.jp/cDNA/>
 MINIVAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
 Kawagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, U.,
 Shikata, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
 Ohneda, E., Yahagi, W., Suzuki, K., Ii, C., Ohtsuki, K., Shibiki, T. and
 Yamamoto, M.
 RFLP Genome Sequencing & Analysis Group: Otsomo, Y., Iida, Y.,
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawanata, M., Kobayashi, M.,
 Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Mura, J.,
 Mizuno, K., Naitkawa, R., Nishura, Y., Oka, M., Ryu, R., Sugano, S.,
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
 Yoshimura, A., Matsubara, K. and Murakami, K.
 Genesome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi, Y., Aizawa, K.,
 Akimura, T., Arkawa, T., Carlini, P., Fukuda, S., Hanaagaki, T.,
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,

FEATURES

Source

Koyas, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinsagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takehashi, F., Takaku-Akchira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

ORIGIN

Query Match	15.2%;	Score 221.6;	DB 8;	Length 2148;
Best Local Similarity	55.9%;	Pred. No. 1.6e-31;		
Matches 529; Conservative	0;	Mismatches 399;	Indels 18;	Gaps 5

QY	20	GGCCACAGACTTGGATGGGGGCTTCGCAACGGCCCTACAGATCGAAGGCGCGCTCA	79
Db	203	TCCCGAGAGATTATCTTTCGGACCGGCTCCGGGCTTATCAGTACGAAGGCGCTGCA	262
QY	80	AAGAAAGTGCGCCGCGCCCTCCATCTGGGACAGTACTGCACTGGAGCCATCGCGCA	139
Db	263	ACGAAGGGGGGAGAGGGCCGAGCATCTGGGACACGTACCTCACATTCAGGCAAAATTG	322
QY	140	CCAACGGCGCCAACGGGATGTGGCTTGGGATACACACCGCATACATGAGACTTGG	199
Db	323	AAGATGGTACAGGGGGATGTAGCACTGGACTTTTACATGCTTACAGAGGATCGA	382
QY	200	ATCTCTTGACCAAGTACGGCGCAAGGCGTACCGCTTCTCTTGTGCTGATGCGCATCA	259
Db	383	ACTTCGTGACCGACATGACATGAGAGCGCTTCGCTTCTTCATTTGCTTGGACAGAGATCC	442
QY	260	TTCCTTCGCGGCGGATCGGATCCGCTCAACAGAGAGGAAATTGATTTTACAGCAAC	319
Db	443	TGCCAAATGMAACATCATGTAAGAGAAATCAACAAGAGGGGATTCCTTCTTCAACACCC	502
QY	320	TGATTTGACGCCCTGTTGAGGCGGGGATATCAOGCTTGGGTGACTTTGATCACTGGGATC	379
Db	503	TGATTTAACGAGGTATATCTTAGAGGGTTGAGGCATTTTGTACCATTTTCAATTCGACA	562
QY	380	TGCCTCAGGGCGCTTACAGATCGCTATGAGAGGCTGAGCTCAACGTGAGAGGATCCAGCTGG	439
Db	563	CCCACAGGCTCTGGAGAGCAAAATACCGGAGTCTCTCA--GTAAAAACATTTGTGAAGG	619
QY	440	ACTTTGAGCGGTATGCGAGGTTGTGCTTTGAAACGTTTGGGGACCGAATCCAGATTCGAA	499
Db	620	ATTTTCGTGACTATGCGGACGTGTGCTTCGTGAGTTCCGCGCACCGGGGTGAAGTATCGA	679
QY	500	TCACCATCAACGAMCCCTGATTCAGCGCATCTATGAGATATGCCACGGGACGACAGCCG	559
Db	680	ACAGGTTCAACGAGCCGATGATCTTTGGCGCGCGGCTTACGGCTTCGGACACCAAGGCC	739
QY	560	CGGCGAGAGC-----AGCATTTAACAAAGACTCCACGAGGGCAACACTGCCACTGAGC	613
Db	740	CCGGCGCTCTCGCGCTACGTCTCCAGAAATGCGCCCGGGCGACTTCGGGACAAGAGC	799
QY	614	CTGTGGCTCTGTGAAAAGGCCCAATATATAGCAATGCCCGCGCGCTGCTTACAGCA	673
Db	800	CGTACGTGCGCGGCAACAATGTGTCTCGCCACGAGGCGGAGGCCCTTCGCTTACCGCC	859
QY	674	GGGACTTTGCGCCCTCGCAAAAAGGGCAAGTGGGATCTCGTCTCAAGGGGACATCAATG	733
Db	860	AAAAGTACAGGAGCGCAAAAGGGGCAATGGCATACCAAGGTTGCGACTGTGTCG	919
QY	734	AGCCCTGGGACAGCAATGAGCTTCGGGACAAAGAGCTGTGAGCGACGATGGAATTTCC	793

Db	920	TGCCCT---ACAGGAGAGCGCGCGCCGACAAAGACGCGGTGAGGCGAGCGCTCGACTTCA	976
Oy	794	ACATTGCGTGTTCGCAATCCCATCTTTTGAAAGAGAACTATTCAGAGACATGAAGA	853
Db	977	TGTACGGATGTTTATGATGACCCGATCGTGT---CGGCGACTACCGCGGACATGTGAAGA	1033
Oy	854	AGCAGCTGGGGGAGAGCGCTTCAGCGCCCTCACTCCCGGGGACTTTGCCATCTCTCATATGCCG	913
Db	1034	AGTTGTGCGGAGCGCGCTGCGCAAAATTCACAGCCGACAGTCGAGTTGGTGA---GG	1090
Oy	914	GAGAGACCGACTTTACGCGCATGTGATTTACTACATCTCCAGTTCGCG	959
Db	1091	GATCTACGACTTTCATCGCGCTCAACTCTACACACCACTACCC	1136
RESULT 13			
AK120430			
LOCUS			
DEFINITION	AK120430	2148 bp mRNA linear	PLN 29-OCT-2003
ACCESSION	AK120430		
VERSION	AK120430.1	GI:37990053	
KEYWORDS	FLU CDNA; CAP trapper.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE			
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			
101			
102			
103			
104			
105			
106			
107			
108			
109			
110			
111			
112			
113			
114			
115			
116			
117			
118			
119			
120			
121			
122			
123			
124			
125			
126			
127			
128			
129			
130			
131			

FEATURES	source
ORIGIN	<p>Query Match 15.2%; Score 221.6; DB 8; Length 2148; Best Local Similarity 55.9%; Pred. No. 1.6e-31; Matches 529; Conservative 0; Mismatches 399; Indels 18; Gaps 5.</p>
FEATURES	<p>location/Qualifiers</p> <p>1..2148</p> <p>/organism="Oryza sativa (japonica cultivar-group)" /mol_type="mRNA" /culivar="Nippondare" /db_xref="taxon:39947" /clone="J013095M12"</p>
REFERENCE	<p>JOURNAL COLLECTION, mapping, and annotation of 28K full-length cDNA clones from japonica rice 3 (bases 1 to 2148) Kikuchi, S.</p>
AUTHORS	<p>Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007) This clone is one of the 32K full-length cDNA clones from japonica rice.</p>
JOURNAL	<p>URL: http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project_Team:_Kikuchi,S.,_Saitoh,K.,_Negata,T.,_Kawagashira,N.,_Doi,K.,_Kishimoto,N.,_Yazaki,J.,_Ishikawa,M.,_Yamada,H.,_Ooka,H.,_Hotta,I.,_Kojima,K.,_Namiki,T.,_Ohneda,E.,_Yahagi,W.,_Suzuki,K.,_Li,C.,_Ohtsuki,K.,_Shishiki,T.,_Yanamoto,M._and_Nakahama,Y._FAS_Genome_Sequencing_&_Analysis_Group:_Otsuo,Y.,_Iida,Y.,_Fujimura,T.,_Ikeda,R.,_Ishibiki,J.,_Kawamata,M.,_Kobayashi,M.,_Kodama,T.,_Kurosaki,T.,_Kusunegi,T.,_Lu,M.,_Masuda,H.,_Mura,J.,_Mizuno,K.,_Narikawa,R.,_Nikura,J.,_Oka,M.,_Ryu,R.,_Sugano,S.,_Sugiyama,A.,_Suzuki,Y.,_Tanoda,Y.,_Ueda,M.,_Xie,Q.,_Yokozaki,S.,_Yoshimura,A.,_Matsubara,K._and_Mutakami,K._Genome_Exploration_Research_Group_in_Riken_Genomic_Sciences_Center_and_Genome_Science_Laboratory_in_Riken:_Adachi,J.,_Aizawa,K.,_Akamura,T.,_Arikawa,T.,_Carninci,P.,_Fukuda,S.,_Hangaki,T.,_Hara,A.,_Haseizume,W.,_Hayashida,K.,_Hayatsu,N.,_Hiramoto,K.,_Hirao,T.,_Hori,F.,_Iida,J.,_Imamura,K.,_Imotoi,K.,_Ishii,Y.,_Itoh,M.,_Kagawa,I.,_Kanagawa,S.,_Kato,H.,_Kawai,J.,_Kishikawa-Hirozane,T.,_Kojima,Y.,_Kondo,S.,_Konno,H.,_Kouda,M.,_Koya,S.,_Kurihara,C.,_Matsuyama,T.,_Miyazaki,A.,_Murata,M.,_Nakamura,M.,_Nishi,K.,_Nomura,K.,_Numasaki,R.,_Ohno,M.,_Osato,N.,_Ore,Y.,_Saitoh,H.,_Sakai,C.,_Sakai,K.,_Sakazume,N.,_Sano,H.,_Sasaki,D.,_Sato,K.,_Shibata,K.,_Shinagawa,A.,_Shiraki,T.,_Sogabe,Y.,_Tagami,M.,_Tagami-Takeda,Y.,_Tagawa,A.,_Takahashi,F.,_Takahashi-Akashi,S.,_Tanaka,T.,_Tomaru,A.,_Toya,T.,_Waki,K.,_Yaenishi,A._and_Hayashizaki,Y._</p>
COMMENT	

Query Match 14.8%; Score 215.8; DB 8; Length 1811;
 Best Local Similarity 51.3%; Pred. No. 2e-30;
 Matches 714; Conservative 1; Mismatches 648; Indels 30; Gaps 8;

QY 20 TGGCCACGACCTTGAATGGGGCTTGCACACGGCCGCTTACAGATGGAAGGGCGCTCA 79
 DB 135 TCCCAAGGACCTTCATCTTCGGGCAAGGTTGACAGCTTACGTATAGAGGGCGCTTACA 194
 QY 80 AAGAAGGTGGCGGGCGGCTCCATCTGGGACACGTACCTGCGACCTGAGCGCATCGGCA 139
 DB 195 AAGAAGGGGGCAAGGCTTACAGGCTTGGGACAACTTCACTCACTTCCAGTAAATTT 254
 QY 140 CCAACGGCGCCAAACGGCGATGTGGCTTGCATCACTCAACCGCTACGATAGAGACTTTG 199
 DB 255 TAAACAATGATTAACGGGATGTGGCAATGACTTCTATCAACCATACAAAGAGATGTGA 314
 QY 200 ATCTCTTACCAAGTACGGGCGCAAGGCTTACCGCTTCTCTGTGTGTGTGGATCA 259
 DB 315 GCTCTCTGAAGGACATGACATGATGCTTCCGGTCTTCCATTTGCGTGAACAGATCC 374
 QY 260 TTCCCTCGGGGAGGCTGATCCGCTCAACGAGGAGGAAATTGATTTTACAGCAAC 319
 DB 375 TCCCAATGATCTTGTAGTGAAGAAATTAACAAAGAGGCTTCTTCTACACAGCT 434
 QY 320 TGAATGACGCTTGTAGGCGGGGTATCAACGCTTGGGTGACTTTGTACACTGGGATC 379
 DB 435 TGATCAATGATGTATGACAAAGAGGATGATCCATTTGTCACTATCTTCACTGGGCA 494
 QY 380 TGCCTCAGCGCTTCAAGATCGCTATGAGGCTGGCTCAACGTGAAGAGTCCAGCTGG 439
 DB 495 CCCCCTTGGCTTGAAGAAAGAAATACGAGGATTTCTCA---GTGAAGACATAGTGAAG 551
 QY 440 ACTTTGAGCGGTATGCGAGGTGTGCTTGAACCTTTTGGGACCGGACTCAAGACTGGA 499
 DB 552 AATACGTGACTTTCGCGAGGAGTGTCTTTCGCACTTTCGCGACCGGTCAAGATCTGGA 611
 QY 500 TCAACATCAACGACCTTGAATTCAGGCACTTATGATATGCAACCGGCGACCAACGCC 559
 DB 612 CCACTTTATGAGCCATTCACATATAGCCCTTACGCTACGCAAGGGCGTGTTCGCC 671
 QY 560 CGGGGAGGACGACGCA-----TTAACAAGCACTTCCACGAGGCGACACTGCGCTAGC 613
 DB 672 CGGAGACGATCTCTTCATATGTTTCCAAATCATGCGCGCTGCGGATCTCCAGTGCAGC 731
 QY 614 CGTGGCTCGGTGAAAGCCCAATCATGAGCCATGCCCCCGCTGCGCTTACAGCA 673
 DB 732 CTAACCTTCGGGACACCACTCACTCCACGCTGCGCTGTCCAGCTTTACCGCA 791
 QY 674 GGAAGCTTTCGCCCTCGCAAAAGGGCCAGATCGGATCTGCTCAACGGCGACTACTATG 733
 DB 792 CCAAGTACAGCCCAACACAGAGGACAGATCGGCACTGTGTGTGTACCACTCGCTTGC 851
 QY 734 AGCCTTGGGACAGCAATGAGCTTTCGAGCAAGAGGCTGTGACGACGAGATGGAAATTC 793
 DB 852 TGGCGTACGACAACTCCGACGCT---GACGCTGGCGCTGTGCAACGAGCCTTGAATTCA 908
 QY 794 ACAATGGCTGTGTTTGGCAATCCCATCTTCTTGAAGAAAGACTATCCAGAGACTTGAAG 853
 DB 909 TCTATGGGTGTTCATGAGACCTTA--TGTGTGATGTGTGACTACCCAGGACCACTGAG 965
 QY 854 AGAGAGTGGGAGAGGCTTTCAGACCTTCACTCCGCGGACTTTGCAATCTCAATCCG 913
 DB 966 GTTGGCTCGGCAATCGGCTGCGAAGTTCAAGCTTGAACAGTGGGCAATGTGTAGAGGCT 1025
 QY 914 GAGAGACGACTTCTTACGGAATGATTAACATCCAGTTTCGCGGCACTTACG 973
 DB 1026 CTTA---GCACTTCATCGGAGTTAATTACTAACACCTACTAGGATTAACCGC 1082
 QY 974 GTCCCGTCCCGGACG-----ACTATCTCGGCGCATTCATGAGACCAAGAGATA 1027
 DB 1083 CGCTTAACTCCAAAGACTTATCTTACGACTTCAACCGCGCCAAACACCGCGCTTCC 1142
 QY 1028 AGGACGGACCCCGCTTGGGAGAGAGCGGCTCGCTGCTGCTGCTCGCCGAGCA 1087

DB 1143 GTAATGCAAAACCCATCGGTCCACAGGAATTTACACGATCTTCTCACTACCTCCAG 1202
 QY 1088 TGTTCGGGAAGCATCTTCGCCCGGCTGTACGCGCTGTACGCGCAAGCCCATTAATCACCG 1147
 DB 1203 GTCTCGGTAGCTTCCTCTTACCAAGAGAGAAATACAAACCCGA---CCATATATG 1259
 QY 1148 AAGAAGGATCCCGTGCCTTGGAGAGAGACATGACGTGCGAGAGGCGGTCAACGACC 1207
 DB 1260 TTACAGAAAAGGCACTGATGAGGTATACAGACACATCGCCAGAGGCGCTCAAGATG 1319
 QY 1208 CTTTCGCACTCCGTACTTTGACTCGCACTTGGACTGATTTTCCAAAGCCATTACCCAG 1267
 DB 1320 GACACAGGATCGAGTTTCACTCAAGGACCTGCACTTCCGCAACACGCGCAT---CAAG 1376
 QY 1268 ACCGCTCTCTTCACAGGGGTACTTTGCGTGGCGCTTGTCTGATTAATTGATGTGAC 1327
 DB 1377 ATGGGTGAACGTGAAGGGCTTACTTACCGTGACATTGATGACTGCTTGTGAGTGGGTG 1436
 QY 1328 ATGCTACGGAACCAATTTGCGGTCAAGTTTCAAGACTTACACACCTCAAGCGCACGC 1387
 DB 1437 ACGGCTACCTTGGACAGGTTTGGCTTCACTGCTCAACGCAAGACGCTTACAGGCTTACC 1496
 QY 1388 CCAAGAGTCTGC 1400
 DB 1497 GCAAGAGTCAAGC 1509

RESULT 15
 AX652977 1503 bp DNA linear PAT 22-MAR-2003
 LOCUS AX652977 Sequence 2847 from Patent WO03000898.
 DEFINITION AX652977
 ACCESSION AX652977
 VERSION AX652977.1 GI:29155791
 KEYWORDS
 SOURCE
 Oryza sativa
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; Oryza.
 REFERENCE
 1 Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
 Katagiri,F., Qian,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
 Plant genes involved in defense against pathogens
 JOURNAL Patent: WO 03000898-A 2847 03-JAN-2003;
 FEATURES
 location/Qualifiers
 1..1503
 /organism="Oryza sativa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:4530"

ORIGIN

Query Match 14.7%; Score 214.6; DB 6; Length 1503;
 Best Local Similarity 51.0%; Pred. No. 3.4e-30;
 Matches 720; Conservative 1; Mismatches 660; Indels 30; Gaps 8;

QY 20 TGGCCACGACCTTGAATGGGGCTTGCACACGGCCGCTTACAGATGGAAGGGCGCTCA 79
 DB 95 TCCCAAGGACCTTCATCTTCGGGCAAGGTTGACAGCTTATCAGTATGAGGGCGCTTACA 154
 QY 80 AAGAAGGTGGCGGGCGGCTCCATCTGGGACAGTATGCGCACTGAGGACCATCGGCA 139
 DB 155 AAGAAGGGGGCAAGGCTTACAGGCTTGGGACAACTTCACTCACTTCCAGTAAATTT 214
 QY 140 CCAACGGCGCCAAACGGCGATGTGGCTTGCATCACTCAACCGCTACGATGAGAGACTTTG 199
 DB 215 TAAACAATGATTAACGGGATGTGGCAATGACTTCTATACCGATACAGAGAGATGTGA 274
 QY 200 ATCTCTTACCAAGTACGGGCGCAAGGCTTACCGCTTCTCTTGTGTGTGTGGATCA 259
 DB 275 GCTCTCTGAAGGACATGAACATGATGCTTTCGGGTCTTCCATGATGTGTGACCAAGATCC 334

QY 260 TTCCCTCGCGGAGGCTGATCCCGTCAACAGAGGGAATTGATTTACAGCAAC 319
Db 335 TGCCAAATGGATCCTTATGAGAGGAATAAACAAGAGGGGCTTCTTCAACAAGCT 394
QY 320 TGAATGACGCCCTGTGAGGCGGGTATCAAGCCTTGAGTCTTGTACCTGGGATC 379
Db 395 TGATCAATGATGATAGCAAAAGGAGTATCCATTGTTCATCTTTCATCGGAGCA 454
QY 380 TGCCTCAGGCGCTTCAAGATGCTATGAGGCTGGCTCAAAGTGAAGAGGCTCAGCTGG 439
Db 455 CCCCCTTGGCTCTGGAAGAAGAAATACGAGGATTCCTCA--GTGAAGACATAGTGAAG 511
QY 440 ACTTTAGCGGTATGCGAGGTTGCTTTGAACGTTTTGGGGACCGAGTCCAGAACTGA 499
Db 512 AATAGGTGACTTTCGCGAGGCTGCTTTCGAGATTGCGGACCGTGTCAAGTCTGA 571
QY 500 TCACATCAAGAGGCCGATTCAGGGCATCTATGATATGCAACGCGCAGCAACGCC 559
Db 572 CCACTTTAATGAGCATTTCATATAGCGCTACGCTACGCGTACGAGGGCGTTTGGCG 631
QY 560 CGGCGAGGAGCA--TTAACAACACTCCACCGAGGCAACCTGCCACTGAGC 613
Db 632 CGGAGCATGCTCTTATATGTTTCAATATGCTATGCGGCTGCTGACTCCAGTCCAGTCCAG 691
QY 614 CGTGCTCGCTGGAAGGCCAGATCATGAGCCATGCCGCGCTGGCGCTTACAGCA 673
Db 692 CTTACCTGTCGACACACATCCACCTTCCAGCTGCGCTGTCCAGCTTACCGCA 751
QY 674 GGGACTTTGCGCCCTCGCAAAAGGCGCAGATCGGCTCTGCTCAACGCGCACTATATG 733
Db 752 CCAAGTACAGCAACAAGAGGAGCAGATCGGCTATGATGATGATGATGATGATGATGATG 811
QY 734 AGCCCTGGAGCAGATGAGCTCGGAGCAGAGAGCTGCTGAGGAGAGGATGAAATTTC 793
Db 812 TGCCGTACGACCACTCCGAGCT--GACCGTGGCGCTGTGCAAGAGAGCTTACCTTCA 868
QY 794 ACATGAGCTGTTTGGCAATCCCATCTTCTTGAAGAAGGATATCCAGAGCATGAAGA 853
Db 869 TCTATGAGTGTTCATGACCTTA--TCTGCTATGATGATGATGATGATGATGATGATG 925
QY 854 AGCAGCTGGGCGAGAGGCTTCAAGCCCTGACCTCCGCGGACTTTGCCATCTCAATGCC 913
Db 926 GTTGGCTGGCATGCGCTGCGCAGAGTCAAGCTGAGCAGTCCGAGTGGTGAAGGCT 985
QY 914 GAGAGCCGACTTCTACGCGATGAATTAACAATCCAGTTGCGGCGCACTAGAGC 973
Db 986 CCTA---CGACTTCACTGAGTTAATTAATTAACAACAATTAACAATTAACAATTAACA 1042
QY 974 GTCCGTCGCCGA-----GACGAGTATCTCGGCGCATTCATGAGCAACAGAGATA 1027
Db 1043 CGCTTAATCTCAAGAGCTATCTTAAGACCTGAGCAACCGCGCAACACCGGCTTCC 1102
QY 1028 AGGAGCGCAGCCCGTTGCGAGAGAGAGCGGCTCGCTGCGCTCTGCGCGGACA 1087
Db 1103 GTATGAGCAAAACCATGAGTCCACAGAAATTTACACGATCTTCTTAACCTACCTCCAG 1162
QY 1088 TGTTCGAGAGCAATCGCCGCGGTATAGGCGCTGTACGCGAGCCATCTACATCACCG 1147
Db 1163 GTCTCGTGAAGCTCTCTTCAACAAGAGAGATTAACAACAACCGA---CCATCTATG 1219
QY 1148 AGAAGGATGCCGTGCTCGAGAGAGAAATGAGTGAAGAGCGGCTGACGACC 1207
Db 1220 TTACAGAAAAGGAGATGATGAGGATTAACAACAGCACTGCGAGGCGCTCAAGAGATG 1279
QY 1208 CTTTCGCAATCCGTAATTTGACTGCGACTTGGACTGATTTCCAGGCGCATTAACGAG 1267
Db 1280 GACACAGGATCGAGTTCACTCAAAAGCACTGCGAGTTGTCAACACGCGAT---CAAGA 1336
QY 1268 ACCGCGCTGCTGCAAGGAGTCTTGGGTGGGCTGCTGCAATACCTTGAATGTGAG 1327
Db 1337 ATGGAGTGAAGTGAAGGCTACTTCAAGTGAATTCATGAGCTGTTGAGTGGGCTG 1396
QY 1328 ATGGCTAGGACCAAGATTGGCGCTCAAGTTCAAGACTACCACTCAAGCGACGC 1387

Db 1397 ACGGCTTACCTTGAAGAGGTTCCGCTCATCTACGTGACCGCAAGAGCTCAAGCGCTAC 1456

QY 1388 CCAAGAGTCTGCCCTGCTGCTCAAGAGCAT 1418

Db 1457 GCAGAGTCCAGCTATCTGATCGAAGACTT 1487

Search completed: March 28, 2004, 20:52:13
Job time : 5767.87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2004, 19:13:22 ; Search time 576.342 Seconds
(without alignments)
10724.755 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455
Sequence: 1 atgcgcgcagtcgcctcctc.....gggttaagtcgcgcataa 1455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001s:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002s:.*
7: Geneseqn2003as:.*
8: Geneseqn2003bs:.*
9: Geneseqn2003cs:.*
10: Geneseqn2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1454.2	99.9	1455	9	ADD22904 T. reesei
2	1440.6	99.0	1990	9	ADD22902 Full leng
3	598	41.1	735	3	Aafl14935 Trichoder
4	266.6	18.3	1933	3	Aafl14882 Trichoder
5	246	16.9	1404	7	ADA70881 Rice gene
6	214.6	14.7	1503	7	ADA69524 Rice gene
7	205.4	14.1	575	7	ABZ53130 Aspergill
8	182.8	12.6	1734	6	ABZ13477 Arabidops
9	182.8	12.6	1734	7	ADA67833 Arabidops
10	178.6	12.3	1533	4	AAS21370 Human cdn
11	178.2	12.2	2016	4	AAS21370 Human cdn
12	178.2	12.2	2016	6	ABL68245 Human PRO
13	178.2	12.2	2016	6	ABL95734 Human ang
14	178.2	12.2	2016	7	ACD23979 Novel hum
15	178.2	12.2	2016	7	ACA67120 CDNA enco
16	178.2	12.2	2016	7	ACA67120 CDNA enco
17	178.2	12.2	2016	7	ABX89267 DNA enco
18	178.2	12.2	2016	7	ACD41921 Human sec
19	178.2	12.2	2016	7	ACA04150 Human cdn
20	178.2	12.2	2016	8	ADA45772 Novel hum
21	178.2	12.2	2016	8	ADA76203 Human PRO
22	178.2	12.2	2016	8	ADA18853 Human PRO
23	178.2	12.2	2016	8	ADA61476 Homo sapi

ALIGNMENTS

24	178.2	12.2	2016	8	ADB19261	Adb19261 Novel hum
25	178.2	12.2	2016	8	ADB27802	Adb27802 CDNA enco
26	178.2	12.2	2016	8	ADA66281	Ada66281 Novel hum
27	178.2	12.2	2016	8	ADB15845	Adb15845 Human PRO
28	178.2	12.2	2016	8	ADA47631	Ada47631 Human PRO
29	178.2	12.2	2016	8	ADA67426	Ada67426 Human PRO
30	178.2	12.2	2016	8	ADB30433	Adb30433 CDNA enco
31	178.2	12.2	2016	8	ADA85729	Ada85729 Novel hum
32	178.2	12.2	2016	8	ADA96941	Ada96941 Human PRO
33	178.2	12.2	2016	8	ADA79245	Ada79245 Human PRO
34	178.2	12.2	2016	8	ADA67384	Ada67384 Novel hum
35	178.2	12.2	2016	8	ADB16586	Adb16586 Human PRO
36	178.2	12.2	2016	8	ADA91678	Ada91678 Novel hum
37	178.2	12.2	2016	8	ADB14741	Adb14741 Human PRO
38	178.2	12.2	2016	8	ADB18702	Adb18702 Novel hum
39	178.2	12.2	2016	8	ADA93917	Ada93917 Human PRO
40	178.2	12.2	2016	8	ADB19813	Adb19813 Novel hum
41	178.2	12.2	2016	8	ADB13125	Adb13125 Human PRO
42	178.2	12.2	2016	8	ACD98550	AcD98550 Novel hum
43	178.2	12.2	2016	8	ADA74379	Ada74379 Human PRO
44	178.2	12.2	2016	8	ADB24612	Adb24612 Human PRO
45	178.2	12.2	2016	8	ADA82136	Ada82136 Human PRO

RESULT 1	ADD22904	standard; CDNA; 1455 BP.
ID	ADD22904	
XX		
AC	ADD22904;	
XX		
DT	15-JUN-2004	(first entry)
XX		
DE	T. reesei bgl5 beta-glucosidase CDNA.	
XX		
KM	BGL5; beta-glucosidase; fungus; ethanol; sugar;	
KW	fermentative microorganism; endoglucanase; cellobiohydrolase; wine making;	
KW	aroma; detergent; softening agent; cotton; wood pulp; gene; ss.	
XX		
OS	Hypocrea jecorina.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..1455
FT		/tag= a
FT		/product= "BGL5 beta galactosidase"
FT		/transl_except= (pos:37..39,aa:Gln)
FT		/transl_except= (pos:511..513,aa:Xaa)
FT		/transl_except= (pos:1219..1221,aa:Pro)
FT		/note= "Xaa= unknown"
XX		
PN	US2003114330-A1.	
XX		
PD	19-JUN-2003.	
XX		
PF	18-DEC-2001; 2001US-00026140.	
XX		
PR	18-DEC-2001; 2001US-00026140.	
XX		
PA	(DUNN/) DUNN-COLEMAN N.	
PA	(GOED/) GOEDEGEBOUR F.	
PA	(WARD/) WARD M.	
PA	(YAO/) YAO J.	
PI	Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;	
XX		
DR	WPI: 2003-810909/76.	
DR	P-PDB; ADD22903.	
XX		
PT	New substantially purified BGL5 polypeptide with the biological activity	
PT	of beta-glucosidase, useful for degrading biomasses to ethanol.	

PS Claim 2; Page 16-17; 21bp; English.

XX The invention discloses a substantially purified BGL5 polypeptide with
CC the biological activity of a beta-glucosidase. Also claimed is an
CC isolated polynucleotide derived from a fungal source, which comprises a
CC nucleotide sequence encoding an enzyme having beta-glucosidase activity,
CC a Trichoderma reesei bgl5 nucleic acid coding sequence, or its
CC complement, an expression construct including the T. reesei bgl5 cDNA
CC sequence, a purified recombinant enzyme having beta-glucosidase activity,
CC a recombinant host cell comprising a deletion or insertion or other
CC alteration in the bgl5 gene which inactivates the gene and prevents BGL5
CC polypeptide production, an antisense oligonucleotide, a heterologous
CC composition comprising the polypeptide and expressing a heterologous
CC polypeptide having beta-glucosidase activity in an Aspergillus sp. BGL5
CC is useful for producing ethanol, which involves contacting a biomass
CC composition with an enzymatic composition comprising the polypeptide to
CC yield a sugar solution, adding to the sugar solution a fermentative
CC microorganism and culturing the fermentative microorganism under
CC conditions sufficient to produce ethanol, where the biomass composition
CC may be optionally pretreated. The method further involves the addition of
CC at least one endoglucanase or cellobiohydrolase. The pretreatment is with
CC a dilute acid. BGL5 is useful in wine making for enhancing the potential
CC aroma of the finished wine product. The detergent is useful as softening
CC agent and also useful for improving the feel of cotton fabrics and for
CC degrading wood pulp into sugars. The sequence presented is the T. reesei
CC bgl5 cDNA.

XX Sequence 1455 BP; 312 A; 444 C; 423 G; 274 T; 0 U; 2 Other;

Query Match 99.9%; Score 1454.2; DB 9; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCAGTGTGAGTCTGCTCCCAAGCTTTGAATGGGGCTTCGCAACGGCCGCTTAC 60
DB 1 ATGCCAGTGTGAGTCTGCTCCCAAGCTTTGAATGGGGCTTCGCAACGGCCGCTTAC 60
QY 61 CAGATCGAAGCGCGCTCAAGAAAGTGGCGCGCGCGCTTCATCTGGGACACGTATGCG 120
DB 61 CAGATCGAAGCGCGCTCAAGAAAGTGGCGCGCGCGCTTCATCTGGGACACGTATGCG 120
QY 121 CACCTGAGAGCATGCGCACCAACGCGCGCAAGTGGCTTGGATCACTTACAC 180
DB 121 CACCTGAGAGCATGCGCACCAACGCGCGCAAGTGGCTTGGATCACTTACAC 180
QY 121 CACCTGAGAGCATGCGCACCAACGCGCGCAAGTGGCTTGGATCACTTACAC 180
DB 121 CACCTGAGAGCATGCGCACCAACGCGCGCAAGTGGCTTGGATCACTTACAC 180
QY 181 CGTAGAGTGAAGCATTTGATCTCTTGAACCAAGTACGCGCAAGGCTTACCGCTTCC 240
DB 181 CGTAGAGTGAAGCATTTGATCTCTTGAACCAAGTACGCGCAAGGCTTACCGCTTCC 240
QY 241 TTGTGCTGAGTGGCGGATCATTCCTCGCGCGGAGGCTGATCCCGTCAACGAGAGGGA 300
DB 241 TTGTGCTGAGTGGCGGATCATTCCTCGCGCGGAGGCTGATCCCGTCAACGAGAGGGA 300
QY 301 ATTGAATTTCAGCAAACTGATTTAGCGCCCTGTTAGGCGGGGTATCAAGCTTTGGGTG 360
DB 301 ATTGAATTTCAGCAAACTGATTTAGCGCCCTGTTAGGCGGGGTATCAAGCTTTGGGTG 360
QY 301 ATTGAATTTCAGCAAACTGATTTAGCGCCCTGTTAGGCGGGGTATCAAGCTTTGGGTG 360
DB 301 ATTGAATTTCAGCAAACTGATTTAGCGCCCTGTTAGGCGGGGTATCAAGCTTTGGGTG 360
QY 361 ACTTTGATCACTGAGATCTGCTCAAGCGCTTCAAGATGCTTGAAGGCTGCTCAAC 420
DB 361 ACTTTGATCACTGAGATCTGCTCAAGCGCTTCAAGATGCTTGAAGGCTGCTCAAC 420
QY 421 GTGAGAGAGTTCAGCTGAGCTTTGAGCGGTATCGAGGTTGCTTTGAAGCTTTTGGG 480
DB 421 GTGAGAGAGTTCAGCTGAGCTTTGAGCGGTATCGAGGTTGCTTTGAAGCTTTTGGG 480
QY 481 GACCGAGTCCAGAACTGATCACTCAACGAGCCCTTGATTCAGGCGCATATATATAT 540
DB 481 GACCGAGTCCAGAACTGATCACTCAACGAGCCCTTGATTCAGGCGCATATATATAT 540
QY 541 GCGACCGGAGCAAGCGCCCGGAGAGGAGCACTTAAACAAGCACTCCACGAGGGGCAAC 600
DB 541 GCGACCGGAGCAAGCGCCCGGAGAGGAGCACTTAAACAAGCACTCCACGAGGGGCAAC 600

QY 601 ACTGCCACTGAGCGGTGCTCGCTGGAAGAGCCCAAGTATCTAGCCATGCCCGCGCTG 660
DB 601 ACTGCCACTGAGCGGTGCTCGCTGGAAGAGCCCAAGTATCTAGCCATGCCCGCGCTG 660
QY 661 GCCGTTCACAGCAGGAGCTTTGCGCCCTCGCAAAAGGCCAGATCGGATCTGCTCAAC 720
DB 661 GCCGTTCACAGCAGGAGCTTTGCGCCCTCGCAAAAGGCCAGATCGGATCTGCTCAAC 720
QY 721 GGGGACTACTATAGAGCCCTGAGGACAGAAATGACCTCGGACAAAGAGGCTGAGGCGA 780
DB 721 GGGGACTACTATAGAGCCCTGAGGACAGAAATGACCTCGGACAAAGAGGCTGAGGCGA 780
QY 781 CGGATGGAATTTACATTTGCTGTTTGCCTCAATCCCATCTTTTGAAGAAGACTATCCA 840
DB 781 CGGATGGAATTTACATTTGCTGTTTGCCTCAATCCCATCTTTTGAAGAAGACTATCCA 840
QY 841 GAGAGCATGAAGAGCAGCTGGGCGAGAGGCTTCAGCCCTCACTCCCGGAGCTTTGCC 900
DB 841 GAGAGCATGAAGAGCAGCTGGGCGAGAGGCTTCAGCCCTCACTCCCGGAGCTTTGCC 900
QY 901 ATCCCTAATGCGGAGAGACCGAATTTCAAGGATGAATTAATCAATCCAGTTCCGCG 960
DB 901 ATCCCTAATGCGGAGAGACCGAATTTCAAGGATGAATTAATCAATCCAGTTCCGCG 960
QY 961 CGCCACTAGACGGTCCCGTCCCGCGAGACGGAATCTCGGCGCCCATTCATGAGACAG 1020
DB 961 CGCCACTAGACGGTCCCGTCCCGCGAGACGGAATCTCGGCGCCCATTCATGAGACAG 1020
QY 1021 GAGATTAAGAGCGAGCGAGCCCGCTTGGCGAGAGAGCGGCTCGCTGCTGCTCTGCG 1080
DB 1021 GAGATTAAGAGCGAGCGAGCCCGCTTGGCGAGAGAGCGGCTCGCTGCTGCTCTGCG 1080
QY 1081 CCGGACATGTTCCGGAAGCATCTGCGCGGGGTGTAAGGCTGTAAGGCAAGCCATCTAC 1140
DB 1081 CCGGACATGTTCCGGAAGCATCTGCGCGGGGTGTAAGGCTGTAAGGCAAGCCATCTAC 1140
QY 1141 ATCAGCAGAAAGGATGCGCGTCCGCTGAGAGAGAAACATGACGTGCGAGAGCGCTGC 1200
DB 1141 ATCAGCAGAAAGGATGCGCGTCCGCTGAGAGAGAAACATGACGTGCGAGAGCGCTGC 1200
QY 1201 AACGACCCCTTCCGATCCGATCTTGAATCTGCACTTGGATCTGATTTCAAGGCGATT 1260
DB 1201 AACGACCCCTTCCGATCCGATCTTGAATCTGCACTTGGATCTGATTTCAAGGCGATT 1260
QY 1261 ACCGAGAGCGGCGTGGTGTCAAGGGGTACTTGGCGGCGGTTGCTGCAATCTTGGAA 1320
DB 1261 ACCGAGAGCGGCGTGGTGTCAAGGGGTACTTGGCGGCGGTTGCTGCAATCTTGGAA 1320
QY 1321 TGGTCAAGTGGCTTACGAGACCCAGATTCGAGGTCAAGTTCACAGTCAACCTTCAAG 1380
DB 1321 TGGTCAAGTGGCTTACGAGACCCAGATTCGAGGTCAAGTTCACAGTCAACCTTCAAG 1380
QY 1381 CGCAGCGCCCAAGAAATGCTGCTGCTTCAAGAGCAATGTTTGGCGCGGAGAGGTT 1440
DB 1381 CGCAGCGCCCAAGAAATGCTGCTGCTTCAAGAGCAATGTTTGGCGCGGAGAGGTT 1440
QY 1441 AAAAGTGGCGCATTA 1455
DB 1441 AAAAGTGGCGCATTA 1455

RESULT 2
ADD22902
ID ADD22902 standard; cDNA; 1990 BP.
XX
AC ADD22902;
DT 15-JAN-2004 (first entry)
XX
DE Full length T. reesei bgl5 beta-glucosidase cDNA.
XX BGL5; beta-glucosidase; fungus; ethanol; sugar;
XX fermentative microorganism; endoglucanase; cellobiohydrolase; wine making;
KW

KW aroma; detergent; softening agent; cotton; wood pulp; gene; ss.
 XX Hypocrea jecorina.
 XX
 FH Key Location/Qualifiers
 FT CDS 62..1516
 FT /tag= a
 FT /product= "BGL5 beta galactosidase"
 FT /transl_except= (pos:98..100,aa:Gln)
 FT /transl_except= (pos:551..553,aa:Gln)
 FT /transl_except= (pos:572..574,aa:Xaa)
 FT /transl_except= (pos:1226..1227,aa:Pro)
 FT /note= "This codon has an apparent 1 nucleotide deletion
 FT which alters the reading frame. Xaa= unknown."
 XX
 PN US003114330-A1.
 PD 19-JUN-2003.
 PF 18-DEC-2001; 2001US-00026140.
 PR 18-DEC-2001; 2001US-00026140.
 PA (DUNN/) DUNN-COLEMAN N.
 PA (GOED/) GOEDEGEBOUR F.
 PA (WARD/) WARD M.
 PA (YAO/) YAO J.
 PI Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;
 DR WPI; 2003-810909/76.
 DR P-Psdb; ADD22903.
 XX
 PT New substantially purified BGL5 polypeptide with the biological activity
 PT of beta-glucosidase, useful for degrading biomass to ethanol.
 XX
 PS Example 1; Fig 1; 21pp; English.
 XX
 CC The invention discloses a substantially purified BGL5 polypeptide with
 CC the biological activity of a beta-glucosidase. Also claimed is an
 CC isolated polynucleotide derived from a fungal source, which comprises a
 CC nucleotide sequence encoding an enzyme having beta-glucosidase activity,
 CC a Trichoderma reesei bgl5 nucleic acid coding sequence, or its
 CC complement, an expression construct including the T. reesei bgl5 cDNA
 CC sequence, a purified recombinant enzyme having beta-glucosidase activity,
 CC a recombinant host cell comprising a deletion or insertion or other
 CC alteration in the bgl5 gene which inactivates the gene and prevents BGL5
 CC polypeptide production, an antisense oligonucleotide, a detergent
 CC composition comprising the polypeptide and expressing a heterologous
 CC polypeptide having beta-glucosidase activity in an Aspergillus sp. BGL5
 CC is useful for producing ethanol, which involves contacting a biomass
 CC composition with an enzymatic composition comprising the polypeptide to
 CC yield a sugar solution, adding to the sugar solution a fermentative
 CC microorganism and culturing the fermentative microorganism under
 CC conditions sufficient to produce ethanol, where the biomass composition
 CC may be optionally pretreated. The method further involves the addition of
 CC at least one endoglucanase or cellobiohydrolase. The pretreatment is with
 CC a dilute acid. BGL5 is useful in wine making for enhancing the potential
 CC aroma of the finished wine product. The detergent is useful as softening
 CC agent and also useful for improving the feel of cotton fabrics and for
 CC degrading wood pulp into sugars. The sequence presented is the full
 CC length T. reesei bgl5 cDNA.
 XX
 SQ Sequence 1990 BP; 439 A; 611 C; 504 G; 434 T; 0 U; 2 Other;
 QY
 QY Query Match 99.0%; Score 1440.6; DB 9; Length 1990;
 QY Best Local Similarity 99.9%; Pred. No. 0;
 QY Matches 1453; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1 ATGCCGAGTGGCTGCTGCGCCACGACGATTTGATGAGGCTTGGCAAGCGCGCCTAC 60
 QY 62 ATGCCGAGTGGCTGCTGCGCCACGACGATTTGATGAGGCTTGGCAAGCGCGCCTAC 121

QY 61 CAGATGAAGGCGCCCTCAAAAGAGTGCCCGCGCCCTCCATCTGGGACACGTACTGC 120
 Db 122 CAGATGAAGGCGCCCTCAAAAGAGTGCCCGCGCCCTCCATCTGGGACACGTACTGC 181
 QY 121 CACCTGGAGCCATCGGGACCAAGCGCGCCCAACGGGATGATGGCTTGGATCACTACAC 180
 Db 182 CACCTGGAGCCATCGGGACCAAGCGCGCCCAACGGGATGATGGCTTGGATCACTACAC 241
 QY 181 CCGTAGAGTAGAGACTTTGATCTTTGACCAAGTACGGCGCAAAAGGCTTACGCTTCTCC 240
 Db 242 CCGTAGAGTAGAGACTTTGATCTTTGACCAAGTACGGCGCAAAAGGCTTACGCTTCTCC 301
 QY 241 TTGTGCTGTGCTGGGATCAATTCCTCCCTGGCGGAGCTGATCCGCTCAACGAGAGGA 300
 Db 302 TTGTGCTGTGCTGGGATCAATTCCTCCCTGGCGGAGCTGATCCGCTCAACGAGAGGA 361
 QY 301 ATTGAGTTTATACAGCAAACTGATGACGCGCTTGAAGCGGGGTATCAAGCTTGGGTG 360
 Db 362 ATTGAGTTTATACAGCAAACTGATGACGCGCTTGAAGCGGGGTATCAAGCTTGGGTG 421
 QY 361 ACTTTGATCACTGGGATCTGCTCAGGCGCTTCAAGATCGTATGAGAGGCTTCAAC 420
 Db 422 ACTTTGATCACTGGGATCTGCTCAGGCGCTTCAAGATCGTATGAGAGGCTTCAAC 481
 QY 421 GTGGAAGAGCTCAGCTGACCTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGG 480
 Db 482 GTGGAAGAGCTCAGCTGACCTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGG 541
 QY 481 GACCGATCCAGAACTGATCAACATCAACGAACTTGAAGCTTGAAGCTTGAAGTAT 540
 Db 542 GACCGATCCAGAACTGATCAACATCAACGAACTTGAAGCTTGAAGCTTGAAGTAT 601
 QY 541 GCCACCGGAGAGCAAGCGCCCGGAGAGAGAGATTAACAGGATCCACCGAGGCAAC 600
 Db 602 GCCACCGGAGAGCAAGCGCCCGGAGAGAGAGATTAACAGGATCCACCGAGGCAAC 661
 QY 601 ACTGCCACTGAGCCGTGCTGTGGAAAGGCCAGATCATGAGCCATGCGCGCTG 660
 Db 662 ACTGCCACTGAGCCGTGCTGTGGAAAGGCCAGATCATGAGCCATGCGCGCTG 721
 QY 661 GCCGTCTACAGCAGGAGCTTTGCGCCCTGCAAAAGGCGCAATCGGCATCTGCTCAC 720
 Db 722 GCCGTCTACAGCAGGAGCTTTGCGCCCTGCAAAAGGCGCAATCGGCATCTGCTCAC 781
 QY 721 GGCGACTACATGAGCGCTGGGACAGCAATGAGCTCGGGAACAAGAGGCTGTGACGA 780
 Db 782 GGCGACTACATGAGCGCTGGGACAGCAATGAGCTCGGGAACAAGAGGCTGTGACGA 841
 QY 781 CGGATGAATTTCACTTGGCTGTGGCAATCCCATCTTCTTGAAGAAGACTATCCA 840
 Db 842 CGGATGAATTTCACTTGGCTGTGGCAATCCCATCTTCTTGAAGAAGACTATCCA 901
 QY 841 GAGAGCATGAAGAGAGAGCTGGGCGAGAGGCTTCCAGCCCTCCTCCGGGACTTTGCC 900
 Db 902 GAGAGCATGAAGAGAGAGCTGGGCGAGAGGCTTCCAGCCCTCCTCCGGGACTTTGCC 961
 QY 901 ATTCCTTAATCCGGAAGAGCCGACTTCTACGAGTAAATTAATCAACATCCAGTTGCGG 960
 Db 962 ATTCCTTAATCCGGAAGAGCCGACTTCTACGAGTAAATTAATCAACATCCAGTTGCGG 1021
 QY 961 CGGCACCTAGAGCGGTCCCTCCCGAGAGGAGTATCTCGGCGCATATGAGACCAAG 1020
 Db 1022 CGGCACCTAGAGCGGTCCCTCCCGAGAGGAGTATCTCGGCGCATATGAGACCAAG 1081
 QY 1021 GAGAATPAAGAGCGGACGCGCTTGGCGAGAGAGGCGCTCGCTGCGCTCTCTGC 1080
 Db 1082 GAGAATPAAGAGCGGACGCGCTTGGCGAGAGAGGCGCTCGCTGCGCTCTCTGC 1141
 QY 1081 CCGGAGATGTTCCGGAAGAGATCTCGCGCGGGGTATAGGCGCTGTAAGGCAAGCCATCTAC 1140
 Db 1142 CCGGAGATGTTCCGGAAGAGATCTCGCGCGGGGTATAGGCGCTGTAAGGCAAGCCATCTAC 1201
 QY 1141 ATCAACGAGAACGAGATGCCCTGCTGGAAGAGAGAAATGACATGACGTGAGAGAGCGCTC 1200

Db	1202	ATCCACCGAAGACGATGCCCGT-CCTGGAAGAGAAACATGACTGGAGAGAGCCGTC	1260
OY	1201	AACGACCCCTTCCGCATCCGTACTTTGACTCGCATTTGACTCGATTTCGAAGGCCATT	1260
Db	1261	AACGACCCCTTCCGCATCCGTACTTTGACTCGCATTTGACTCGATTTCGAAGGCCATT	1320
OY	1261	ACCCAGAGCGGCGTCTGCTCAAGGGGTACTTTGCTGGGGGCTTCTCGATTAACCTTGGAA	1320
Db	1321	ACCCAGAGCGGCGTCTGCTCAAGGGGTACTTTGCTGGGGGCTTCTCGATTAACCTTGGAA	1380
OY	1321	TGCTCAGATGAGCTACGAGCCAGATTGGGCGTCAAGTTACAGACTACACCAACCCTCAAG	1380
Db	1381	TGCTCAGATGAGCTACGAGCCAGATTGGGCGTCAAGTTACAGACTACACCAACCCTCAAG	1440
OY	1381	CGCAGCGCCAGAGAGTCTGCCCTGGTCTCTCAAGGACATGTTTGGCGCCCGGACAGAGGTT	1440
Db	1441	CGCAGCGCCAGAGAGTCTGCCCTGGTCTCTCAAGGACATGTTTGGCGCCCGGACAGAGGTT	1500
OY	1441	AAAGTGGCGGCATTA	1455
Db	1501	AAAGTGGCGGCATTA	1515

RESULT 3
AAFI4935
ID AAF14935 standard; cDNA; 735 BP.

DT	15-SEP-2003	(revised)
DT	13-MAR-2001	(first entry)
XX		
DE	Trichoderma reesei	EST SEQ ID NO:7458.

The present invention describes a method for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. The method uses fluorescence-labeled nucleic acids isolated from the FF cells and a substrate of expressed sequence tags (EST). The ESTs are used in the methods for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. Monitoring the global expression of genes from FF cells allows the production potential

of the microorganisms to be improved. New genes may be discovered, and possible functions of unknown open reading frames can be identified and gene copy number variation and stability can be monitored. The expression of genes can be used to study how *SF* cells adapt to changes in culture conditions, environmental stress, spore morphogenesis, recombination, metabolic or catabolic pathway engineering. Using ESTs provides several advantages over genomic or random cDNA clones including elimination of redundancy as one spot on an array equals one gene or open reading frame, and organisation of the microarrays based on function of the gene products to facilitate analysis of the results. AAP07478 to AAP11247 represents ESTs from *Fusarium venenatum*; AAP11248 to AAP11853 represents ESTs from *Aspergillus niger*; AAP11854 to AAP14678 represents ESTs from *Aspergillus oryzae*, and AAP14679 to AAP15337 represents ESTs from *Trichoderma reesei*, which are all specifically claimed in the present invention. (Updated on 15-SEP-2003 to standardise OS field)

	Sequence	735 BP;	168 A;	210 C;	205 G;	148 T;	0 U;	4 Other;	
Query Match	41.1%;	Score 598;	DB 3;	Length 735;					
Best Local Similarity	97.0%;	Fred.	No.	6.1e-133;					
Matches	650;	Conservative	1;	Mismatches	13;	Indels	6;	Gaps	4

RESULT 4
AAFL4882

ID AAF14882 standard; cDNA; 1933 BP.
 XX AAF14882;
 AC
 XX
 DT 15-SEP-2003 (revised)
 DT 13-MAR-2001 (first entry)
 XX
 DE Trichoderma reesei EST SEQ ID NO:7405.
 XX
 KM Multiple gene expression; filamentous fungal cell; EST;
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Hypocrea jecorina.
 PN WO200056762-A2.
 PD 28-SEP-2000.
 PF 22-MAR-2000; 2000MO-US007781.
 PR 22-MAR-1999; 99US-00273623.
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
 DR WPI; 2000-594572/56.
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags.
 XX
 PS Claim 89; Page 2989; 3161bp; English.
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring the
 CC global expression of genes from FF cells allows the production potential
 CC of the microorganisms to be improved. New genes may be discovered,
 CC possible functions of unknown open reading frames can be identified and
 CC gene copy number variation and stability can be monitored. The expression
 CC of genes can be used to study how FF cells adapt to changes in culture
 CC conditions, environmental stress, spore morphogenesis, recombination,
 CC metabolic or catabolic pathway engineering. Using ESTs provides several
 CC advantages over genomic or random cDNA clones including elimination of
 CC redundancy as one spot on an array equals one gene or open reading frame,
 CC and organization of the microarrays based on function of the gene
 CC products to facilitate analysis of the results. AAF07478 to AAF11247
 CC represents ESTs from Fusarium venenatum; AAF11248 to AAF1153 represents
 CC ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from
 CC Aspergillus oryzae; and AAF14879 to AAF1537 represents ESTs from
 CC Trichoderma reesei, which are all specifically claimed in the present
 CC invention. (Updated on 15-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 1933 BP; 467 A; 536 C; 568 G; 343 T; 0 U; 19 Other;
 Query Match 18.3%; Score 266.6; DB 3; Length 1933;
 Best Local Similarity 56.2%; Pred. No. 1.1e-53;
 Matches 770; Conservative 1; Mismatches 540; Indels 59; Gaps 13;
 QY 20 TGGCCACGACGCTTTGAATGGGCGCTTGCACACGCGCGCTTACGATGGAAGGCGCGCTCA 79
 DB 250 TGGCCAAAGGACTTTCAGTGGGGGTTCCGACGCGCTGCTACGATGAGAGGCGCGCTCG 309

QY 80 AAGAGGTGGCCGCGCGCCCTTCAT--CTGGACACGTACTGC-----CACTGGAGCCA 132
 DB 310 ACCAGGACGGCGCGCGGCCCGCCAGCATTTCTGGGACACGTTCTGGGCGCCGCGGCAAGA 369
 QY 133 TCGCGACCAACGCGCGCCAAAGCGGATGTGGCTTGGGATCATCAACCGCTACGATGA- 191
 DB 370 TCGCGACGCGCTTCTGCGGCGCTGACGCGGTGCCCACTTCGTAACAACCGCGCGCGGAN 429
 QY 192 GGACTTTGATCTTGTACCAAGTAAGGCGCAAGGCGCTACCGCTTCTGCTGTGGTGC 251
 DB 430 GGAATTTGCGCTCTCAATGCTCTCGGAGCCAAAGAGTACGATTCATCTCTGCTGCTC 489
 QY 252 GCGGATCATTTCCCTCGCGCGCGAGGCTGTGATCCCGTCAACGAGAGGGAATGATTTTA 311
 DB 490 GCGCATCATCTCCGAGAGGCGGAGCGCGGATGCGTMAACGAGCGCGGATGACCACTA 549
 QY 312 GAGCAAACTGATTTAGCGCCCTGTTAGAGCGGAGTATCAAGCTTGGGTGACTTTGACCA 371
 DB 550 GGTCAAGTTCTGTCGAGGACCTGCTCGACGCGGATCATCAGCCCTTCATCACCTCTTCCA 609
 QY 372 CTGGGATCTGCTCAGGCGCTTACAGATGCTATGAGAGGCTGCTCAAGTGAAGAGT 431
 DB 610 CTGGGACCTGCGCGAGGCGCTGATGATGCGGATGCGGAGGCTGTGAACCGCACCGAGTT 669
 QY 432 CCAGCTGACCTTTAGCGGATATCGAGGTTGTCTTGAACGTTTGGGAGCCGAGTCCA 491
 DB 670 CCGGCTGACTTTGAAACCTACGCCCCG--CGTATGTTACGCGCGCTTGGCCAAAGTGGC 728
 QY 492 GAACTGATACCATCAACGAGCAGGCTGATTCAGGCGCATATGATATGACACCGGACG 551
 DB 729 CAATGATACCATCAACGAGCAGGCTGATTCAGGCGCATATGATATGACACCGGACG 788
 QY 552 CAAGCGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 611
 DB 789 CTTCGCGCGCGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 818
 QY 612 GCGGTGCTGCTGTAAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 671
 DB 819 GCGGTGAGCGTGTGGCGCAACATCTCTGTCGCGCCACGCGCGCGCTGACAGGCGTACCG 878
 QY 672 GAGGACCTT--TGGCCCTTGGCAAAAGGCGCGAGATCGGATTCGCTCAACGCGGACTA 728
 DB 879 CGACGACTTCAACCGCTT 938
 QY 729 CTATGAGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 788
 DB 939 CACTTACCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 988
 QY 789 ATTTCACATTTGCTGTTGCTTTCGCAATCCATCTTCTTGAAGAGGAGGAGGAGGAGGAG 848
 DB 999 GTTCTTACGCGCTGCTGTTGCGGATCCATCTTCTTGGGAGACTTACCGGCGTCCA--TGC 1057
 QY 849 GAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 908
 DB 1058 GCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1115
 QY 909 TGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 968
 DB 1116 CCAAGGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1175
 QY 969 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1028
 DB 1176 AGCTCG--CCGCTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1233
 QY 1029 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1086
 DB 1234 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1292
 QY 1087 ATGTTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1146
 DB 1293 GATTTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1352
 QY 1147 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1199

Db 1353 GACGGANACCGNACGAAGCATTTAAGGGGAGAGGACCTTGGCAAGAGAAAGATTTNT 1412
Qy 1200 CAACGACCCCTTCCGCATCCGCTACTTTGACTCGACATTCGAT-TTCCAGGCC 1257
Db 1413 CGAATATGATCTTCAAGGCTCAAGTACTATACAAATACATCCGTCATGGGATACCCGCCG 1472
Qy 1258 ATTACCGAGACCGGCGTGGTCAAGGGGATCTTGGTGGGCGGTCTGATTAATTTG 1317
Db 1473 TGGAGCTGGACGGGCTCAAAAGCTCAAAAGGATCTTGGCTGGTGGCTCATGGACAACTTT 1532
Qy 1318 GAATGTCAGATGGCTACGAGCCAGATTCGCGCTCAAGTTCAAGACTA 1367
Db 1533 GAGGGGCGGACGCGCTACGTAATAAGGTTTGGGGTTACTTAAAGGATTA 1582

RESULT 5
ADA70881
ID ADA70881 standard; DNA; 1404 BP.
XX
AC ADA70881;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 4204.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN MO203000898-A1.
PD 03-JAN-2003.
XX
PE 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
PI Katsagiri F, Ques S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 4204; 899bp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 1404 BP; 356 A; 385 C; 396 G; 266 T; 0 U; 1 Other;

Query Match 16.9%; Score 246; DB 7; Length 1404;
Best Local Similarity 54.5%; Pred. No. 8.7e-49;
Matches 725; Conservative 0; Mismatches 565; Indels 40; Gaps 10;

Qy 62 AGATCGAAGCGCGCTCAAGAGAGTGGCGCGCGCGCTCATCTGGACACGATCGCC 121
Db 86 AGATCGAAGCGCGCTCAAGAGAGTGGCGCGCGCGCTCATCTGGACACGATCGCC 145

Qy 122 ACTGAGCCATGCGCAGCAACGACGCGCCCAAGCGGATGTCCTTGATCTACTACACC 181
Db 146 ACACCTTTGAAAGATATCCAGACTTGCAGCAATCTGATGTTGAGATGATCAGTACACC 205
Qy 182 GCTACGATGAGGACTTTGATCTCTTGAACCAAGTACGCGCGCAAGGCTTACCGCTTCCCT 241
Db 206 GTTTCGAGAGGATATACAGCTATGACAGACATGGGATGATGCGTATCGGTTCTCGA 265
Qy 242 TGTGTGTCGCGGATATATTCCTCCGCGCGGAGGCTGATATCCCTTCAACGAGAGGAA 301
Db 266 TAGATGTGTCAGAAATATACCAATGATG------TTGGTCAAGTCAATCAAGCTGTGA 319
Qy 302 TTGAGTTTACAGCAAACTGATGAGCCCTGTTGAGCGGGGTATCAAGCTTGGGTGA 361
Db 320 TCGACCACTACCAACAGCTGATGATGACATTTTACCAAAAGAAATTCAGCCATATGTGA 379
Qy 362 CTTTGTACCACTGGGATTCGCTCAGGCGCTTCAGCATGCTATGAGGCTGCTCAAG 421
Db 380 CACTTACCACTGGGATTCGCTCAGGCGCTTCAGCATGCTATGAGGCTGCTCAAG 439
Qy 422 TGAAGAAGTCCAGCTGAGCTTGAACGGTATGCGAGTGTGCTTTGAACGTTTGGG 481
Db 440 GG---CAGATAGTGAACGATTTGCGCGCTGACGCGAGAGTGTGCTTCAAGGAGTTGGG 496
Qy 482 ACGAGTCCAGAACTGATTCACATCAACGATCCCTGATTCAGGCTATGATATG 541
Db 497 ACAGGTGAAGCACTGATTCACATCAACGATCCCTGATTCAGGCTATGATATG 556
Qy 542 CCAACGCGAGCAACGCGCGCGGAGAGGAGCATTA-----ACAAGCACTCCACCGAG 595
Db 557 ACGCAGGCGCTCCAGGCGCGCGCGCTGCTCCTGTGCTTCACTCTTACAGGCGG 616
Qy 596 GCAACACTGCTGACTGAGCGGTGCTGCTGAAAGGCCAGATCATGAGCCATGCGCGG 655
Db 617 GCAACTCCGACACGAGCCCTACGTCGTGCGCCACCACTTACCTCGCCACGCGCGG 676
Qy 656 CCGTGGCGCTTACACAGGAGCTTTCGCGCGCGGAAAGGCGGAGTCCGATTCGCG 715
Db 677 CGCGCAGCACTTACAGGAGCAAAATACAGGCGGAGGAGGCGGCTTGGATAGCT 736
Qy 716 TCAACGCGCACTATGAGCCCTGAGCAGCAATGAGCTCGGAGCAAGAGGCTGCTG 775
Db 737 TCGAGCTGATGTCGTCGAGCCGATGTCACACACGAT---CGACATCGAGGGGCGCA 793
Qy 776 AGCGAGGATGGAATTTCACTTGGCTGTTTCCCAATCCCATCTTCTTGAAGAGAT 835
Db 794 AGAGAGCGGAGAGTTTCACTGATGATGTTGCTGATCCGTTCTT---CGCGACT 850
Qy 836 ATCCAGAGCAATGAAGAAGCAGTGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCT 895
Db 851 ACCCGCGACGATGAGGCGGAGGAGGCTGGGAGAGGCTGCCAGGTTCAAGCGGATGAG 910
Qy 896 TTGCACTCTCAATGCGGAGAGACCGACTTCTACGCGATGAATTAACATCCAGT 955
Db 911 CGCGCGCTCTCAAG---GGGCGCTGATTTGTCGGGCTTAACACATCAACATCACT 967
Qy 956 TGGCGGCACTAGACGCTCCGCTCCGAGAA-----CGACTATTCGCGGCTATCA 1010
Db 968 ACACGAGGCAACAAACACCAACATCATCGGACATTTGCTCAACAAACCTTGGAGACA 1027
Qy 1011 TTAGCAGCAGAGAAAT-----AAGACGCGAGCGCGCTTGGCGAGGAGAGCGGCTCG 1063
Db 1028 CCGGACCGCTCAGCCTCCCATTCAGAAATGGAAGCCATTTGAGATAGGCGAAATTTGA 1087
Qy 1064 CCGTGTGCTCTCTGCGCGGACATGTTCCGAAAGCATTCGCGCGGATGATCGGCTGT 1123
Db 1088 TATGCTGTATCATTTGTCGCCGAGGAGATGAGAGGCTGATGAATATGTAAGAAAGT 1147
Qy 1124 ---ACGGCAAGCCCATCTACATCAACGAGAAAGGATGTCGCCGCTTGAAGAGGAACA 1180
Db 1148 ACACAGCCCAACAGTATCACTGAAACGGGATGATGACGACAAACCCGTTCA 1207

Qy	1181	TGAGGTGGAAGAGCGCGTCAACAGACCCCTCCGATCCGCTACCTTTGACTCGGACCTTG	12480
Db	1208	TTTTCATCAAGAGACGCGCTCAAGACAGCAAGAGATCAAAATTCACATGACTTACCTCA	12675
Qy	1241	ACTCGATTTCCAAAGGCCATTACCCAGACGCGCGTGTGCTCAAGGGGTACTTTGCGTGGG	13000
Db	1268	CCATATCGGCGTCTTTCATCAAGAGAGACGGGTGCAAGTACGTGGGTACTTTCGCTGGT	13272
Qy	1301	CGTTGCTCGATTACTTTGGAA-TGGTCAGATGGCTACGGACCCAGATTTGGCGGTACGTTT	13599
Db	1328	CTTCGCTGGACAACTGGGGAATTTGGCGGCGCGGATCACTCTCGAGATTCGGGCTCTACTTC	13879
Qy	1360	ACAGACTTACA	1369
Db	1388	GTGGACTTACA	1397

RESULT 6
ADA69524
ID ADA69524 standard; DNA; 1503 BP

AC ADA69524 ;

DT 20-NOV-2003 (First entry)

Rice gene, SEQ ID 2847.

KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.

OS *Oryza sativa*.

PN WO2003000898-A1.

PD 03-JAN-2003

PF : 22-JUN-2001; 2001WO-1B001105.

PR 22-JUN-2001; 2001WO-1B001105.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI	Chang H,	Chen W,	Cooper B,	Glazebrook J,	Goff SA,	Hou Y,
PI	Katagiri F,	Quan S,	Tao Y,	Whitham S,	Xie Z,	Zhu T,
PI						Zou G;

DR WPI; 2003-175290/17.

PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.

PS Claim 6; SEQ ID NO 2847; 899pp; English.

CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC the expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.

5Q Sequence 1503 BP; 380 A; 420 C; 379 G; 324 T; 0 U; 0 Other;

Query Match	14.7%;	Score 214.6;	DB 7;	Length 1503;
Best Local Similarity	51.0%;	Pred. No. 2.8e-41;		
Matches 720; Conservative	1;	Mismatches 660;	Indels 30;	Gaps 8;

20 TGCCCAAGCAGCTTTGAATGGGGCTTCGCAAGGGCCGCTTACAGATCGAAGGCCCGTCA 79

Db	95	TTCCCAAGAGCATTCATCTTGCGGACAGGTTACAGAGTTATCAAGTAGAGGCGCC	TA	150
Oy	80	AAGAAGTGGCCGCGCCCGCTCATCTTGGGACAGTATCGCACCTGGAGCCATCGCGCA		139
Db	155	AAGAAGGGGGCAAAAGGCTCTTAGCGCTCTGGGACAACTTCACATTCACAGTAAATTT		214
Oy	140	CCAAAGGCGCAACGGGATGTGGCTTGGCATCACTACCAACCGGTACAGTAGAGGACTTGG		199
Db	215	TAAACAAATGATATACGGGATGTGGCAAAATGACTTATATCCACCGATATCAAGAGATGTGA		274
Oy	260	TTCCCTCGCGGCGAGCTGATCCGTCACACGAGAGGAAATTGATTTTACAGCAAC		319
Db	335	TGCCAAATGATCCTTAGTGAAGAAATAAACAAGAGGGGTGGCTTTCTACAAACGCT		394
Oy	320	TGATTTAGCGCCCTGTATGAGCGGGGTATCAACGCTTGGGTGACTTTGTACACATGGGATC		379
Db	395	TGATCAATGATATGTCATATGCAAAAGGATATCCATTATGACTATTTTCCACTGGGACA		454
Oy	380	TGCTTCAGGCGCTTCACGATGCGTATGAGAGGCTGGCTCAACGTGGAAGAGTCCAGCTGG		439
Db	455	CCCCCTGGCTCTGGAAAGCAATACGAGAGATTCCTCA---GTGAAGACATAGTGAAG		511
Oy	440	ACTTTGACCGGATGCGAGGTGTGCTTTGAAAGTTTTGGGAGCCGAGTCCAGACTGGA		499
Db	512	AATATGATGACTTTCGGGAGGTGTGCTTTGCGGAGTTCCGCGACCGGTATCAAGTATCGGA		571
Oy	500	TCACCATATCAAGAACCTTGATTCAGGCCATCTATGATATGCAACCGGACGAACGCCCC		559
Db	572	CCAAATTTAATGACCCATTGACATATAGCCCTTACGGCTACGGGAAAGGGGTGTTTCGC		631
Oy	560	CGGCGAGAGACAGA-----TTAACAGGCACTCAACGAGGCGACACTGCGCACTGAG		613
Db	632	CGGAGCAATGCTCTTCAATATGTTTCCAAATCATGCGCGTTCGGATCTCCAGTCCGAGC		691
Oy	614	CGTGGCTTCGCTGAAAGGCGCCAGATCATAGACCATGCGCGCGCGTGGAGCGGTCTACAGA		673
Db	692	CCTACCTCGTGGGACACCAATCCACTTCGCCACGCTGCGCGGTCTCAGGTTTATCCGCA		751
Oy	674	GGAAGTTTCGCGCCCTCGGCAAAAGGGCCAGATTCGGCATCTCGCTCAACGGGACCTATCATG		733
Db	752	CCAAATGACAGCCAAACAGAAAGGACAGATTCGGCATGATGGTGTACACCATCGATTCG		811
Oy	734	AGCCCTGGGACAGCAATGAGCGCTCGGGACAAAGAGGCTGCGAGCGACCGATGGAATTTCC		793
Db	812	TGCGGTACGACATCTCGACGCT---GACCGTGGCGCTGTGGCAACGGAAGCTTACATTTCA		868
Oy	794	ACATTGGCTGGTTTGGCCAATCCCATCTTCTTGTGAAGAAAGACTATTCAGAGACATGAGA		853
Db	869	TCTATGGGTGGTTCATGGAACCTTA---TGTGTGACATGGTGAATACCCAGGACCATGAGAG		925
Oy	854	AGCAGCTGGGAGAGGCTTCAGCCCTCACTCCCGGCACTTTGCCATCTCTCATGCCG		913
Db	926	GTTTGGCTCGCAATCGGCTGCGAGAGTTTCAACGCTTGAACAGTCGGGCATGGTGAAGGCT		985
Oy	914	GAGAGACCGAATCTTACAGCATGAATTAATCAACATCCAGTTGCGGCGGCACTTAACG		973
Db	986	CCTA---GCACTTATCGAGGTTAATTAATCTACACCACTTACTAGCTTAAGATATACCGC		1042
Oy	974	GTCCCGTTCGCCGA-----GACGCACTATCTCGGCGCATCATAGACACACAGAGATA		1027
Db	1043	CGCCTAATCCAAACGACGTATCTTAGACACTTCGACAAACCGGCGCAACACACGCGCTTCC		1102
Oy	1028	AGGACGGGAGCCCCGTTGGCGAGAGAGCGGCTCGCTGAGCTGCGCTCTGCTCCGACA		1087
Db	1103	GTAATGCGCAAAACCATGGTTCACAGGAATTTACACCGATCTTTCTTCACTACCTTCAG		1162
Oy	1088	TGTTTCGGAAGCATCTGCGCGCGGTGTATGAGGCGGTATAGGCAAGCCATCTATACACG		1147
Db	1163	GTTCTCGTAGCTCTCTCTTCAACCAAGAGAGATCAACCAACCGCA---CCATCTATG		1219

Qy 1148 AGAAGGATGCCCCGCTGGAGAGAGAAATGACGAGGAGGAGCCGCAACGACC 1207
Db 1220 TTACGAGAAACGCGATTCATGAGGTAACACACGACACTGCCAGAGGCGCTCAAGGATG 1279
Qy 1208 CTTCCGCATCCGCTACTTTGATGCACTTGACCTGCAATTTCCAAAGCCATTACCCAGG 1267
Db 1280 GACACAGGATGATGCTTCAAGCACTGCAAGTTCGTCAACACGCGCAT---CAAGA 1336
Qy 1268 ACGGCGTGTGTCGAAGGGGTACTTTGGTGGGCGCTTCTGATTAATCTTGAATGTAG 1327
Db 1337 ATGGGTTAAGGTGAAGGCTACTTCACTGAGCACTTCACTGATCTTTAAGTGGGTG 1396
Qy 1328 ATGGCTACGAGCCAGATTCGCGCTCACGTTACACAGTACCAACCCCTCAAGGCGACGC 1387
Db 1397 ACGGCTACCTTGACAGGTTCCGCTCATCTACGTGACCGCAAGACGCTCAAGCGCTACC 1456
Qy 1388 CCAAGAGTGTGCTGCTGCTGCTCAAGACAT 1418
Db 1457 GCAAGAGTCCAGCTACTGATCGAAGACTT 1487

RESULT 7

ABZ53130
ID ABZ53130 standard; cDNA; 575 BP.

XX ABZ53130;

DT 28-MAR-2003 (first entry)

DE Aspergillus oryzae polynucleotide SEQ ID NO 2243.

XX Aspergillus oryzae; fermentation; fungus; industrial; EST;

KM expressed sequence tag; gene; ss.

XX Aspergillus oryzae.

OS WO200279476-A1.

FN 10-OCT-2002.

PD 22-MAR-2002; 2002WO-IB000890.

XX 30-MAR-2001; 2001JP-00098371.

PR (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

PA (NARE-) NAT RES INST BREWING.

PA (NORQ) NAT FOOD RES INST MIN AGRIC.

XX Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H,

PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;

XX WPI; 2003-046817/04.

DR Detection of expression of specific Aspergillus genes for monitoring the

PT fermentation and growth conditions of the fungus, using DNA probes.

XX Claim 1; SEQ ID NO 2243; 48bp + Sequence listing; Japanese.

XX The invention relates to a polynucleotide having any of 6006 specific

CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under

CC specific culture conditions including one or more of eutrophic,

CC oligotrophic, solid, early germination, alkaline, high temperature, low

CC temperature or maltose culture or polynucleotides stringently hybridizing

CC to these sequences. The polynucleotides are useful for monitoring the

CC progress of fermentation and the growth conditions of a fungus,

CC especially of Aspergillus oryzae which is widely used in industrial

CC fermentation. Also monitoring for fungal contamination. Note: The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 575 BP; 143 A; 154 C; 134 G; 144 T; 0 U; 0 Other;

Query Match 14.1%; Score 205.4; DB 7; Length 575;
Best Local Similarity 63.0%; Pred. No. 3,4e-39;
Matches 317; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

Qy 17 CTCTGCCCAACGACTTGAATGGGGCTTCCGCAACGCGCGCTTCCAGATGGAAGCGCG 76
Db 71 CTCTCCCAACCGACTTCTATGGGGATTTCCCAACGAGGATTTACGATTTGAAGAGCGG 130
Qy 77 TCAAGAGGTGGCGCGCGCGCGCTTCCATCTGGAACAGTACCTGCACTGGAACCTGCGC 136
Db 131 TAAATGAGGACGAGCGAGCTTCATTCATCTGGGATACATTTGCAAGATCCCGTAAAA 190
Qy 137 GCAACACGCGGCAACGCGGATGGCTTGGATCACTACCAACCGCTAGAGAGACT 196
Db 191 TTGCTGAGGCGCAATGGGATGGCTTGGCACTCAATCAATGCAACACAGAGATA 250
Qy 197 TTGATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTTGTGCTGCGCGA 256
Db 251 TTGCTTGTCTCAAGGCTTGGGCTTGGCCAGGCAATACGTTTCTACTCTCTTGGTCCGCA 310
Qy 257 TCATTCCTTCCGCGCGCGAGCTGATCCGCTCAACGAGAGGAATTGATTTTACAGCA 316
Db 311 TTATTCACCTCGGTGGCGCAAGACCCCATCATGAGAAAGGCTTGCAATATTATATA 370
Qy 317 AACTGATGACGCGCTTGGTGAAGGCGGATATCAAGCTTGGGACTTTGTACCACTGGG 376
Db 371 AGTTCTGACGACCTGCACTGCTGTGTATCACTCTCTCTGTTACTTGTTCACCTGGG 430
Qy 377 ATCTGCTCAGGCGCTTCAAGATGCTATGAGAGCTGCTCAACGAGAGAGTCCAGC 436
Db 431 ATCTTCCGATGAGCTGCAAAAGCGTACGCGGCTCTCTCAATAAGAGATTGCTGG 490
Qy 437 TGAATTTGAGCGGTATGAGAGTGTGCTTTGAACCTTTTGGGAGCCGAGTCCAAACT 496
Db 491 CAGACTTGGCCACTGACGACGCGATGTTTCAAGGCTTTCGGCTCGAAATTAACACT 550
Qy 497 GGATCAACATCAAGAMCCCTGG 519
Db 551 GGATCACTTCAACGAGCCATGG 573

RESULT 8

ABZ13477
ID ABZ13477 standard; DNA; 1734 BP.

XX ABZ13477;

DT 21-JAN-2003 (first entry)

DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1282.

XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX Arabidopsis thaliana.

PN WO200216655-A2.

XX 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US026685.

XX 24-AUG-2000; 2000US-0227866P.

PR 26-JAN-2001; 2001US-0264647P.

PR 22-JUN-2001; 2001US-0300111P.

XX (SCRI) SCRIPPS RES INST.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Harper JF, Krebs J, Wang X, Zhu T;

XX WPI; 2002-304127/34.

PT Identifying a stress condition to which a plant cell has been exposed and
 PT producing plants with increased tolerance to these abiotic stresses.

XX Claim 144; SEQ ID NO 1282; 577pp + Sequence Listing; English.

XX
 CC The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising: (a) contacting nucleic acid
 CC representative of expressed polynucleotides in the plant cell with an
 CC array or probes representative of the plant cell genome; and (b)
 CC detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stresses. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
 CC in methods of the invention. Note: The sequence data for this patent is
 CC not represented in the printed specification but is based on sequence
 CC information supplied to Derwent by the European Patent Office

XX
 SQ Sequence 1734 BP; 556 A; 316 C; 385 G; 477 T; 0 U; 0 Other;

Query Match 12.6%; Score 182.8; DB 6; Length 1734;

Best Local Similarity 49.2%; Pred. No. 1.1e-33; Mismatches 658; Indels 27; Gaps 6;

Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

QY 18 TCTGCCAAGCACTTGAATGGGGCTTCCGACGCGCCCTACCAATGCAAGCGCCGCT 77
 DB TTTTCTGATGATTCATCTTTGGGAGAGCTGCTGGGCTTTCAGATGCAAGAGTGAAC 149
 QY 78 CAAGAAAGTGGCGCGCGCCGCTTCATCTTGGACAGTACTGCCACCTGGAGCCATCGC 137
 DB AAGTGAAGGTGGCGAGCTTCAACTATATGGGATCACTTCAGCCTCAAGTATCAGAAAG 209
 QY 138 CACCAACGCGCGCCGAGATGGCTTGGCTGATCACTACACCGGCTAGAGAGACTT 197
 DB GACCAAAATGCAATATGCAATATGCAATATGATTTTATATCAATGCAAGATGACAT 269
 QY 198 TGATCTCTTGAACAGTACGCGCAAGCGCTACCGCTTCTCTTGTGTGTGCGGAT 257
 DB AAAATGATGAAGAGCACTAAACATGACGCTTCCGATTTTCAATCTCGGTCAAGATT 329
 QY 258 CATTCCTCCGCGCGGAGCGTGTGATCCCGTCAACGAGAGGAAATTGATTTACAGCA 317
 DB AATATCCCACTGGAAGCACTAAAGATGAGTAAACAAAGAGGTGTCAATTTTACAGAG 389
 QY 318 ACTGATTGACGCGCTTGTGAGCGGGGTATACGCGCTGGGTGACTTTGTACCACTGG 377
 DB TCTCATAGCAAACTTCTGCTATATACATACAACTTGCATATACGCTCTATATTTGG 449
 QY 378 TCTGCTCAGCGCTTACAGATGCTATGAGGCTGCTCAACGTGGAAGAGTCCAGCT 437
 DB CCACCCCAATCTTTTGGAGAGCAATATGTGCTTTCTTAAGCCCTTAATTCGTAGA--- 506
 QY 438 GGACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGACCGGAGCTCAAGACTG 497
 DB AGATTTTTCAGATTTTGAAGATTTGTTTGAAGAGTTTGGAGATTAAGTTAGTGTG 566
 QY 498 GATCAACATCAAGCAAGCTGATTCAGGCGCATCTATGATATAGCAACCGCGAGCAAGC 557
 DB GACCAACATCAAGCAAGCTTATATTAATGATGTGGGTTATATGATCAAGAGAGG 626
 QY 558 CCGGAGCAGAGAGC-----AGCATTAACAAGCACTCCACGAGGGAACACTGCCACTGA 611
 DB GGTGAGCATGTGTAAGATGAGTAAAGCAAGAGTGTGAGGCTGAGATTGAGTACCGA 686
 QY 612 GCGGTGCTGCTGTAAGAGCCCAAGATCATGAGCCATGCCGCGCGTGGCGCTTACAG 671
 DB GCTTTTCATTTGTTTACATCACTCTTCTTGGCCATGCGCGCTGAGTGAAGAAATTTG 746
 QY 672 CAGGAGCTTTGCGCCCTCGCAAAAGGCGAGATCGGATCGCTCAAGCGGAGCACTACTA 731
 DB AAAATGTGAAGAAAGCTTGCATGATGAGCCAAATTTGGATGATGATCATACCAAGATGTT 806
 QY 732 TGAGCCCTGGAGCAGCAATGAGCTTCGGGCAAGAGAGCTGTGAGGAGCAGATGGAATT 791

DB 807 CGAGCCTTATCAATTCGATTCACATGACATTAAGAGCAGCTGAACGAGCTTCTCCTT 866
 QY 792 TCACATTGCTGTTGGCAATCCCATCTTCTTGAAGAGAGTATCCAGAGCATGAA 851
 DB 867 TGAATTTGATGGCATCTTGTATCACTTTC---ACGAGATTTTCAAGATTTGTA 923
 QY 852 GAAGCAGTGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTGGCCATCTCAATGC 911
 DB AAGTACGGGGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 983
 QY 912 CGAGAGACCGACTTCAAGCATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 971
 DB TTCA---TCGATTTGTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1040
 QY 972 CCGTCCCGTCCCGAGACGACATATCGGCGCCATCCATGTA-----GCACAGGA 1022
 DB 1041 TCACATGACCCGTAAGAAACCTCGTTTCAAAATGACACCATGATGAAATGAAACTGAC 1100
 QY 1023 GAATAAGACGCGAGCCCGTGGCGAGAGAGCGGCTCGCTGGCTGCTGCTGCCC 1082
 DB 1101 TAAATCAGTGTGATCATCATGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1160
 QY 1083 GGCATGTTCCGAGACATCTGCGCCCGGTGTACGCGCTGTAC--GGCAAGCCATCTA 1139
 DB 1161 GGAAGGCTTACGAAAGTCTTAACTATATTAAGAGAGATCAATATACATGCGCAGCTTA 1220
 QY 1140 CATCACCGAAGACGATGCGCCCGTGGCGAGAGAGAAATCAATGCGAGAGAGCGCT 1199
 DB 1221 CATCAAGAAATATGAATTAATGACAGACGAGGTACAAACCAAGAGAAATTTGT 1280
 QY 1200 CAAGCAGCCCTTCCGATCCGATCTTGTGATCTGCACTTGTGATCTTGTGATCTTGTGAT 1259
 DB 1281 GAAGGACATTTAGATTTGAATGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 1340
 QY 1260 TACCAAGACGCGGTGCTCAAGGGGTACTTGGCTGGGCGTGTGATTAATTTGA 1319
 DB 1341 AGTGAAGATGGGTGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 1400
 QY 1320 ATGTCAGATGCTGACGAGCCAGATTTGG 1349
 DB 1401 ATGAGGACATGATTAACATGCAAGATTTGG 1430

RESULT 9
 ADA67833
 ID ADA67833 standard; DNA; 1734 BP.
 XX
 AC ADA67833;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Arabidopsis thaliana gene, SEQ ID 71.
 XX
 KW Plant; bacterial infection; fungal infection; viral infection; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO2003000898-A1.
 XX
 PD 03-JAN-2003.
 XX
 PF 22-JUN-2001; 2001WO-IB001105.
 XX
 PR 22-JUN-2001; 2001WO-IB001105.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 XX Katsairi F, Qian S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;
 DR WPI; 2003-175290/17.
 XX

PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.

Claim 6; SEQ ID NO 71; 899pp; English.

CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.

Sequence 1734 BP; 556 A; 316 C; 385 G; 477 T; 0 U; 0 Other;

Query Match 12.6%; Score 182.8; DB 7; Length 1734;

Best Local Similarity 49.23; Fied: NO. 1.18-55;
Matches 664: Conservative 1; Mismatches 658

Matches	664;	Conservative	1;	Mismatches	658;	Indels	27;	Gaps	6;
---------	------	--------------	----	------------	------	--------	-----	------	----

OY	18	TCGCGCCCAACACTTTGATGAGGGGCTTGCGCAACGGCGCGCTACCGAGATGAAGGCGCCGT	77
Db	90	TTTTCTGTGATTTTCATCTTTGGGACACCTGCTCTGGCGTTTCAGTACGAAGGTGCAAC	149
OY	78	CAAGAAGGTGCGCGGCGCCGTCCATCTGGGACAGTACTGCCACTGGAGCCATCGCG	137
Db	150	AAGTGAAGGTGGCAAGTCTCCAACTATATGGGATCACTTCACGCTTCACTATCCAGAAAG	209
OY	138	CACCAACGGCGCCACAGCGGATGGGCTTGCGATCACTCAACCGGTATGATGAGACTTT	197
Db	210	GACCAAAATGCAATTAAGCAGATGTACCAATTGATTTTATCATCGTTACAGAGTACAT	269
OY	198	TGATCTCTTGACCAGTAAGGCGCAAAAGGCTACGCGCTTCTCCCTGTGTGTGCGGAT	257
Db	270	AAATGTAGTAAGGAGCTAAACATGACGCTTCCGATTTTCAATCTCGTGTCAAGATT	329
OY	258	CATTCGCCCTGCGCGAGGCTGATATCCGTTCACAGAGAGGAATTGAGTTTACAGCA	317
Db	330	AAATCCAGTGGAAAGCTAAAGGATGAGTAAACAAAGAGGTGCAATTTCAAGAG	389
OY	318	ACTGATTTGACGCCCTGTGTGAGCGGGGTATCAGCCCTTGGTGACTTTGTACCACTGGGA	377
Db	390	TTCTCATAGAGAACTTCTGTCTATATGACATACAAACCTTGATGACGCTCATCATTTGGGA	449
OY	378	TCGTGCTCAGGCGCTTCCAGATGCGTATGAGAGGCTGCTCAAGTGAAGAGGTCCAGCT	437
Db	450	CCACCCCAATCTTTGGAGACCAAAATGTGTGCTTTCTAAGCCCTTAAATCTATGA---	506
OY	438	GGACTTTGACGGGATGCGAGGTTGTGCTTTGAACGTTTGGGAGCCGAGTCCAGAACTTG	497
Db	507	AGATTTTCGAGATTTTGCAGAATTTGTTTGAAGAGTTTGGAGATTAAGTAAAGTGTG	566
OY	498	GATCACCATCAACGAWCCCTTGATTCAGGCACTATGATATATGCCACGGCAGCAACGC	557
Db	567	GACAACTATCAACGAACTTATATATATATGACTGTGGCGGGTTATGATCAAGTAAACAAGGC	626
OY	558	CCGCGGAGAGGAGC-----AGCATTAACAAGCACTCCACGAGGGGCAACCTGCCACTGA	611
Db	627	GGCTGAGCGATGCTCAAAATGGGTAAACGAAAGTGTCAAGCTGAGATTCGAGTACCGA	686
OY	612	GCGCTGAGCTCGCTGAAAGGCCCAAGATCATAGGCAATGCCGCGCGCTGGCCTTACAG	671
Db	687	GCGTTACATTTGTTTCACTACACACTCTTCTTGCCCATGCGCTGCACTGGGAAGATTTTCG	746
OY	672	CAGGACCTTTGCCCCCTGCAAAAGGCGAGATGGGATCTCGGTCAACGGCGCATCTA	733
Db	747	AAATATGTGAAAAAATCTTGCAATGATGGCCAATTTGGATAGTACTTATCACCAAGATGTT	806
OY	732	TGACCCCTGGGACAGCATAGAGCTTCGGGACAAAGAGGCTGCTAGCGACGATGTGAATT	791

Accession	Gene	Species	Length (bp)
D807	CGAGCCTTATCATTCGATTTCAACTGACGATYAAAGAAAGCAGCTGAAGAGACTCTTGCCCTT	866	
Oy	792 TCACATTGGCGGTGTTGCCATCCCATCTTCTTGAAGAAAGACTATCCAGAGACATGAA	851	
D8	867 TGAATTGGATGGCATTTTGATCCAGTCATTC--ACGAGATATATCCAGAGATTGTA	923	
Oy	852 GAAGCAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGCCATCTCCATGC	911	
D8	924 AAATGACGGCGGAAATTAATTAACCTTATTAATTAATTAATTAATTAATTAATTAATTA	983	
Oy	912 CGGAGAGACCGACTCTTAACGGCAGATTAATTAACATCCAGTTCCGCGCCACTTGA	971	
D8	984 TTCA--TCGATTTTCGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1040	
Oy	972 CGGTCCGCTCCCGAGACGAGCTATCTCGGCGCATTCATGA-----GCACGAGGA	1022	
D8	1041 TCACATGACCTCGTAAATACTCGTTTCAATACTGACACCATGATGGAATGGAATCTGAC	1100	
Oy	1023 GAATTAAGACGGCAGCCCGCTTGGCGAGAGAGGCGCTGCTGCTGCGCTCTGACC	1082	
D8	1101 TAATTCAGTGTGTCATCATGAGACCTGGGAAAGAAAGGGCTTCTTATTTTCAACACC	1160	
Oy	1083 GGAACATGTTCCGGAAGCATCTCGCCGGGTGTACGGCCCTGAC--GGCAAGCCCATCTA	1139	
D8	1161 GGAAGCTTACGAAAGTTCTTAATTAATTAATTAAGAGATACATTAATCATGCTGCTA	1220	
Oy	1140 CATCAACGAGAACGATATCCCGTCCCTGGAGAGAGAAATGACGTGCGAGAGGCGCT	1199	
D8	1221 CATCAAAAGAAATGAAATTAATGACAAACGAGAGGTGCAAAACCAAGAAAGAAATTTG	1280	
Oy	1200 CAAGACCCCTCCGATCCGTACTTGTGACTTCGCACTTGGACTCGATTTCCAAAGCCAT	1258	
D8	1281 GAAAGACACATTTTGAATTAATGACAAACGAGAGGTGCAAAACCAAGAAAGAAATTTG	1340	
Oy	1260 TACCCAGAGCGCGCTGCTGCTCAAGGGTACTTTGCCGCGGCGTTCGATTAATTTGGA	1319	
D8	1341 AGTGAAGATGGGTGTGACGTAAAGGATATTAACCATGCTCATGATGAGCAATTTTGA	1400	
Oy	1320 ATGGTCAGATGGCTACCGAGCCAGATTCGG 1349		
D8	1401 ATGGAGCATGATACATCGCAAGATTTGG 1430		
RESULT 10			
ADA70381			
ID	ADA70381 standard; DNA; 1533 BP.		
XX	ADA70381;		
XX			
XX	20-NOV-2003 (first entry)		
XX			
XX	Rice gene, SEQ ID 3704.		
XX			
XX	Plant; bacterial infection; fungal infection; viral infection; rice;		
XX	gene; ds.		
XX			
XX	Oryza sativa.		
XX			
XX	WO2003000898-A1.		
XX			
XX	03-JAN-2003.		
XX			
XX	22-JUN-2001; 2001WO-IB001105.		
XX			
XX	22-JUN-2001; 2001WO-IB001105.		
XX			
XX	(SYGN) SYNGENTA PARTICIPATIONS AG.		
XX			
XX	Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;		
XX	Katagiri F, Qian S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;		
XX	WPI; 2003-175290/17.		

XX Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.

PS Claim 6; SEQ ID NO 3704; 899bp; English.

XX The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC the expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.

XX Sequence 1533 BP; 414 A; 365 C; 402 G; 352 T; 0 U; 0 Other;

Query Match 12.3%; Score 178.6; DB 7; Length 1533;

Best Local Similarity 49.3%; Pred. No. 1.1e-32; Mismatches 694; Indels 30; Gaps 8;

Matches 705; Conservative 0; Mismatches 694; Indels 30; Gaps 8;

12 GCTAGCTTGGCCCAAGCACTTTGAATGGGCTTCGCAACGGCCCTACAGATCGAAG 71
 126 GCGAGCTTCCCGAGGGGCTTCTGTTCCGACGGCCGCTCGCGTACAGGTGAGGG 185
 72 CGCGCTCAAGAGAGTGGCCGCGCCGCTTCATCTGGAGACGTACTCCACTGAGCC 131
 186 CATGGGAGAGAGGGTGGCCGGGCGCTTAGCATCTGGAGCGTTTACAGAAACGAG 245
 132 ATGGCGCACCAAGGGCGCAAGCGATGGCTTGCGATCACTACACCGCTACATGA 191
 246 GAGCATCCCTAATATATGCAACACCGATGACGTTGATGATATATAGTACAGGA 305
 192 GCACTTGAATCTCTTACCAAGTACGCGCAAGCCCTACCGCTTCTCTGTCGTGTC 251
 306 AATGTGACATATATAGAACATATGGGCTTTGATGCTATAGATTTTCCATCTCTGTC 365
 252 GCGGATCATTCCTCGCGCGGAGGCTGATCCGCTCAACGAGAGGAAATGATTTTA 311
 366 AAGAAATTTTCCCAATATGGGAGCTGGGATG-----GTGAACAGAGAAAGAAATGATTATTA 419
 312 CAGCAAACTGATGAGCGCCTGTTGAGCGGGGTATCAAGCCTTGGGTGACTTTGATCA 371
 420 CAACAGATTATATGATTACATGTTAAGAAAGGATCAAAACGTTACGCAAAACCTTACCA 479
 372 CTGGGATCTGCTCAGGCGCTTACAGATCGCTATGAGGCTGAGCTCAACGTGAAGAGT 431
 480 CTATGACTTACCTTATGAGCACTCCATGAGCACTTATGAGGTGCTAAGC---CCAAACAT 536
 432 CCAAGTGAATTTGAGCGGTATGCGAGTGTGCTTTGAAAGTTTGGGAGCCGAGTCCA 491
 537 CGTAGAGGCGTTTGCAGACTACCAAGATTTCTCTTCCAGACGTTGAGAGACGAGTGA 596
 492 GAATCTGATCACTATCAAGCAACCTTGATTCAGGCTATATGATATGCCACCGGAG 551
 597 GGACTGTTTACCTTCAATGAGCGGAGTGGCTGCTCTAGAGATATGACATGGCTT 656
 552 CAACGCGCGGAGAGAGCATTTAACAAGCACTCCACGAGGAGCAACTGCCACTGA 611
 657 CATTGCCCGGAGAGGTCTCTGCTGTATG-----CAGAGGCAACTCCACGACAGA 710
 612 GCGGTGCTTGTGTAAGGCGCAATCATGAGCCATGCCGCGCTGGCGCTTACAG 671
 711 ACCATATCTTGTGCGCAACCATCTCATCTTCTCATGCTGCGCGCTCAAGGATACCG 770
 672 CAGGACTTTGGCCCTCGAAAAGGCGCATCGGATCTCGCTCAACGCGCATCTACTA 731
 771 CGAAAATATCATCTTATCAAAAAGGAGGATTTGAAATTTCTTGGATTTTGTGTGTA 830

QY 732 TGAGCCCTGGGACAGCAATGAGCTTCGGGACAGAGGCTGTGAGCGACGATGAAATT 791
 DB 831 TGAACCATTCAGTACAGCAATGCT---CATAGGCTGACAGCAGAGGCGAAGATTT 887
 QY 792 TCACATTGGCTGTTTGGCAATTCCTATCTTTGAAAGAAAGACTATCCAGAGCATGA 851
 DB 888 TCACCTGGATGTTCTTGAACCCCATATATC---CATGTGGTATCCATATCTGATGCT 944
 QY 852 GAAGCACTGGGCGAGAGGCTTCAGCCCTCACTCCCGGAGACTTTGCCATCTCAATGC 911
 DB 945 CGAAATCTCAGAGAGAGATGCAACCTTCACTAGTATGAGAGTCCAGATGTTGAAAGA 1004
 QY 912 CGAGAGACCGCACTTCAAGCAATGATTAATCAACATCCCACTGCGCGCCACTAGA 971
 DB 1005 CTCATATGATATATGTTGGATCAACACACTCTTTCTATCATAGAAAGACCTGGGCG 1064
 QY 972 CGGTCCGTCCTCCGAGAGCGATATTTGGCGGCATCAATGACACAGAGAAATAGG- 1030
 DB 1065 ATGGAACTGACACCGACAGATTATCAGATGACCTGGCAATGTTGGTTGGCTATGACG 1124
 QY 1031 --ACGGCAGCGCCGTTGGCGAGAGAGCGGCTCGCTGAGCTGCTGCTGCGCGACAT 1088
 DB 1125 AAGCGCTGTCCTCATTTGAGCTCAGCAAACTCTTCTGCTTTACATTTGTCATAGGG 1184
 QY 1089 GTTCCGGAAGCATCTGCGCCGCTGACGCGCTGTACGCGCAAGCCCATCTACATCA 1148
 DB 1185 AATCAACAAAGGCTGTGACCTATATGAAAGAAACATATGAGAAACCTTACATGATCTTTC 1244
 QY 1149 GAAAGATGCGCTGCTGAGAGAGAAACATGACGTGCGAGAGGCGCTCAACGACC 1208
 DB 1245 TGAATAATGATATGAGACC---AACCTGGCAACGTCATATATCACTACAGGTGTGATATAC 1301
 QY 1209 CTTCGCGATCCGCTTATGATCGCATCTGACCTGATTTCCAGGCCATTTACCCAGGA 1268
 DB 1302 AGTAAGATCAATATCT---ACAGAAATCAATCACTAGCTACGCTACAGAGCGATAGCA 1358
 QY 1269 CGCGCTGCTGCTCAAGGAGTACTTTGCGTGGGCTGTGCTGATTAATGATGTGAGA 1328
 DB 1359 TGGTGCCTCAAGTATGATGATCTTTGCTGTGATCTTGTGACAACTTGCAGTGAAGGCT 1418
 QY 1329 TGGCTTACGACCAATTTGGCGCTCACGTTACAGACTACCAACCTCAAGGCGACGCC 1388
 DB 1419 CGGTGACACTTCCGCTTTGGCATGCTTACGTGACTACAGACCTTAAGAGGTATACC 1478
 QY 1389 CAAGAATGTCGCTGCTCCTCAAGGACATGTTTGGGCGCCGCGAGAGG 1437
 DB 1479 CAAGACTAGCTTCTGTTCAAGAACTGCTTCAAGTAAAGAGG 1527

RESULT 11
 AAS21370
 ID AAS21370 standard; cDNA; 2016 BP.
 XX AAS21370;
 AC
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human cDNA sequence encoding for PRO9820 polypeptide.
 XX
 KW Human secretory and transmembrane; PRO; mammalian; cancer; lung; breast;
 KW prostate; cervical; tumour necrosis factor-alpha; TNF-alpha; cartilage;
 KW ear; proliferation; glucose; free fatty acid; skeletal muscle; adipocyte;
 KW A-peptide; factor VIIA; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200140466-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 01-DEC-2000; 2000MO-US032678.
 XX
 PR 01-DEC-1999; 99MO-US028301.

PR 01-DEC-1999; 99WO-US028634.
 PR 02-DEC-1999; 99WO-US028551.
 PR 02-DEC-1999; 99WO-US028564.
 PR 02-DEC-1999; 99WO-US028565.
 PR 09-DEC-1999; 99US-0170262P.
 PR 16-DEC-1999; 99WO-US030095.
 PR 20-DEC-1999; 99WO-US030911.
 PR 20-DEC-1999; 99WO-US030999.
 PR 30-DEC-1999; 99WO-US031243.
 PR 30-DEC-1999; 99WO-US031274.
 PR 05-JAN-2000; 2000WO-US000219.
 PR 06-JAN-2000; 2000WO-US000277.
 PR 06-JAN-2000; 2000WO-US000376.
 PR 11-FEB-2000; 2000WO-US003565.
 PR 18-FEB-2000; 2000WO-US004341.
 PR 18-FEB-2000; 2000WO-US004342.
 PR 22-FEB-2000; 2000WO-US004414.
 PR 24-FEB-2000; 2000WO-US004914.
 PR 24-FEB-2000; 2000WO-US005004.
 PR 01-MAR-2000; 2000WO-US005601.
 PR 02-MAR-2000; 2000WO-US005841.
 PR 03-MAR-2000; 2000US-0187202P.
 PR 10-MAR-2000; 2000WO-US006319.
 PR 15-MAR-2000; 2000WO-US006884.
 PR 20-MAR-2000; 2000WO-US007377.
 PR 21-MAR-2000; 2000WO-US007532.
 PR 30-MAR-2000; 2000WO-US008439.
 PR 17-MAY-2000; 2000WO-US013705.
 PR 22-MAY-2000; 2000WO-US014042.
 PR 30-MAY-2000; 2000WO-US014941.
 PR 02-JUN-2000; 2000WO-US015264.
 PR 05-JUN-2000; 2000US-0209832P.
 PR 28-JUL-2000; 2000WO-US020710.
 PR 11-AUG-2000; 2000WO-US022031.
 PR 23-AUG-2000; 2000WO-US023522.
 PR 24-AUG-2000; 2000WO-US023328.
 PR 08-NOV-2000; 2000WO-US030952.
 PR 10-NOV-2000; 2000WO-US030873.
 XX (GETH) GENENTECH INC.
 PA Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerlitsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart RA, Tumas D, Watanabe CK, Wood WJ, Zhang Z;
 XX WPI; 2001-408281/43.
 DR P-PSDB; AAU12298.
 XX
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing PRO
 PT polypeptides, and detect the presence of mammalian tumors e.g. lung,
 PT breast, prostate, cervical.
 XX
 XX Claim 3; Fig 253; 813pp; English.
 PS
 XX AAS1244-AAS21518 encode for novel human secretory and transmembrane PRO
 CC polypeptides. The PRO polypeptides are useful to detect other PRO
 CC polypeptides, to link bioactive molecules to cells expressing PRO
 CC polypeptides, to modulate biological activities of cells expressing PRO
 CC polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample. Some
 CC of the 275 sequences are also useful to stimulate the release of tumour
 CC necrosis factor-alpha (TNF-alpha) from human blood, the proliferation or
 CC differentiation of chondrocytes, the proliferation or gene expression in
 CC pericyte cells, the release of proteoglycans from cartilage, the
 CC proliferation of inner ear utricular supporting cells or of T-
 CC lymphocytes, the release of a cytokine from peripheral blood monocytes
 CC (PBMCs), or the proliferation of endothelial cells. Some of the PRO
 CC polypeptides may modulate glucose or free fatty acid uptake by skeletal
 CC muscle cells or by adipocytes; or inhibit binding of A-peptide to factor
 CC VIIA. The PRO polypeptides can be used in assays to identify molecules
 CC involved in binding interactions. The polynucleotides encoding PRO

CC polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy
 XX
 SQ Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;
 Query Match 12.2%; Score 178.2; DB 4; Length 2016;
 Best local Similarity 54.5%; Pred. No. 1.5e-32;
 Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;
 OY 15 AGCTGCCCCAAGCTTTGAAATGGGGCTTCCCAACGGCCGCTACAGATGAAAGCCG 74
 DB 237 AACCTTCCTCTTGGCTTCTCTGGGGCTGGGCACTTCTGCTACACAGAGGGCCG 296
 OY 75 CGTCAAGAAGGTGGCCGGCCGCTCCATCTGGGACAGTACTGCACTTGGAGCCATC 134
 DB 297 CTGGAGACAGAGAGGAAAGGGCTGATCTGGGACGCTTACACACAGTGGGAAGG 356
 OY 135 GCGCACCAACGG--CGCAAGCGGATGTGGCTTGGCTACACTACACCGCTAGATGA 191
 DB 357 GAAGTGTCTGGGAATGAGACGGCAGATGTAGCTGTGACGGCTACTACAGGTCCAG 416
 OY 192 GGACTTGATCTGTGACCAAGTACGGCGCAAGGCTACCGCTTCTCTGTGTGTC 251
 DB 417 GGAATCATTTCTGTGAGGAACTGACATGCAACATTCATTCCTCTGTGTGGCC 476
 OY 252 GCGGATCATTTCCCTCGCGGCGAGGCTGATCCGTCACAGAGAGGAATGATTTTA 311
 DB 477 CCGGCTCTGCCACAGGCATCCGACCGGAGGAGTGAACAAGAAATCAATTTCTA 536
 OY 312 CAGCAACTGATGAGCGCCCTGTGAGGGGGGTATCAGGCTTGGTACTTGTACCA 371
 DB 537 CAGTGATCTTATGATGATCCCTTCTGAGCAGCAACATCATCTCCATGTGACTTGA 596
 OY 372 CTGGGATCTGCTCAGGCGCTTCAGATCGTATGAGGCTGAGCTCAAGCTGAAGAGT 431
 DB 597 CTGGGATCTGCCACAGGTGTCTCAGGTCAAAATCGGAGGTGCAAAATGTGAGTAGC 656
 OY 432 CCAAGTGAATTTGAGCGGTATGCGAGTTTGTCTTGAACGTTTTGGGACCGAGTCCA 491
 DB 657 CAA---CTACTTCAGAGACTACCCCAACCTGTCTTTGAGCCCTTTGGGACCGTGA 713
 OY 492 GAATGATCATCATCAACAGAACGCTGAGATTACGCGCATATGATATGATCCACCG 551
 DB 714 GCATGATATCATCTTCACTGATCTCTCGGCAATGCGAAGAAAGGCTATGAGAGGG 773
 OY 552 CAACGCCCGCGGAGAGACAGATTTAACAGCACTCCACGAGGCAACCTGCCACTGA 611
 DB 774 CAAATGGCGCGGGCCTGAAGCTCCGC-----GGCACCGG 806
 OY 612 GCCGTGCTGCTGGAAAGGCCAGATCATGAGCATGCCCGCGGTGGCCGTCTACAG 671
 DB 807 CTTTACAGGACAGACACACATCATTTAAGCCACCAAACTGGCATTTCTTATTA 866
 OY 672 CAGGAGCTTTGCGCCCTCGCAAAAGGCGCAGATGCGCATCTCGCTCAACGCGA 731
 DB 867 CACCACTGGCGGAGCAAGCAAGCATGCTGTGGGAATTTCACTGAATGTGACTGGG 926
 OY 732 TGAACCTGGAGACGAAATGAGCTTGGGACAGAGGCTGTGAGCGAGATGAAAT 791
 DB 927 GGAACCTGTGACATTAATTAACCCCAAGAGACTTGAAGCTGGCCGAGAAATAC 986
 OY 792 TCAATTTGGCTGCTTTTCCCATCCCATCT 820
 DB 987 CTGTCTGGCTGTGTTTCCCAACCCCATTT 1015
 RESULT 12
 ABL88245
 ID ABL88245 standard; cDNA; 2016 BP.
 XX
 XX ABL88245;
 AC
 XX 16-MAY-2002 (first entry)
 DT

Human PRO9820 cDNA sequence SEQ ID NO:347.

Human; angiogenesis; cardiac; cytoskeletal; antiangiogenic; hypotensive;
vulnerable; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
gene therapy; cardiovascular disorder; endothelial disorder; cancer;
angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
age-related macular degeneration; arterial restenosis; angina;
rheumatoid arthritis; myocardial infarction; thrombophlebitis;
lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
wound healing; chromosome mapping; gene mapping; gene; ss.

Homo sapiens.

WO200200690-A2.

03-JAN-2002.

20-JUN-2001; 2001WO-US019692.

23-JUN-2000; 2000US-0213637P.

20-JUL-2000; 2000US-0219556P.

25-JUL-2000; 2000US-0220624P.

28-JUL-2000; 2000WO-US020710.

02-AUG-2000; 2000US-0222695P.

17-AUG-2000; 2000US-00643657.

23-AUG-2000; 2000WO-US023522.

24-AUG-2000; 2000WO-US023328.

07-SEP-2000; 2000US-0230978P.

18-SEP-2000; 2000US-00664610.

24-OCT-2000; 2000US-0242922P.

08-NOV-2000; 2000US-00709238.

08-NOV-2000; 2000WO-US030952.

10-NOV-2000; 2000WO-US030873.

01-DEC-2000; 2000WO-US032678.

20-DEC-2000; 2000US-00742559.

20-DEC-2000; 2000WO-US034956.

22-JAN-2001; 2001US-00767609.

28-FEB-2001; 2001US-00796498.

28-FEB-2001; 2001WO-US006520.

01-MAR-2001; 2001WO-US006666.

09-MAR-2001; 2001US-00802706.

14-MAR-2001; 2001US-00808689.

22-MAR-2001; 2001US-00816744.

05-APR-2001; 2001US-00828366.

10-MAY-2001; 2001US-00854208.

10-MAY-2001; 2001US-00854280.

25-MAY-2001; 2001US-00866028.

25-MAY-2001; 2001WO-US017092.

30-MAY-2001; 2001US-00870574.

30-MAY-2001; 2001WO-US017443.

01-JUN-2001; 2001WO-US017800.

(GETH) GENENTECH INC.

Baker KP, Ferrara N, Gerber H, Gertlisen ME, Goddard A,
Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF,
Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W,
WPI: 2002-090516/12.
P-PSDB; ABB84990.

One hundred and eighty seven nucleic acids encoding PRO polypeptides,
useful in diagnosis and treatment of cardiovascular (e.g. myocardial
infarction), endothelial or angiogenic disorders in a mammal.

Claim 2; Fig 347; 565pp; English.

ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytoskeletal,

antiangiogenic, hypotensive, vulnerable, antiarteriosclerotic
activities, and can be used in gene therapy. The PRO polynucleotides,
proteins, agonists and antagonists are useful for treating or diagnosing
a cardiovascular, endothelial or angiogenic disorder in a mammal, e.g.
cardiac hypertrophy, trauma, cancer, age-related macular degeneration,
atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
angina, myocardial infarction, thrombophlebitis, lymphangitis, tumour
angiogenesis (such as breast carcinoma and liver carcinoma) and wound
healing. The PRO polynucleotides have applications in molecular biology,
including use as hybridisation probes, and in chromosome and gene
mapping. ABB8259 to ABB8267 represent primers and probes used in the
exemplification of the present invention

Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

Query Match 12.2%; Score 178.2; DB 6; Length 2016;
Best Local Similarity 54.5%; Pred. No. 1.5e-32;
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

Qy	15	AGCTGCCCCAAGCACTTGTGAATGGGCTTCGCAAGCGCGCCTTACAGATCGAAGGCGC	74
Db	237	AACTTCCCTCTTGGCTTCTCTGGGCGTGGGCACTTCTGCTTACAGAGAGGCGC	296
Qy	75	CGTCAAGAAAGTGGCGCGCGCTCATCTGGGACAGTACCTGACCTGGAGGCATC	134
Db	297	CTGGGACAGAGAGGAAAGGCGCTAGCATCTGGGAGCTTTCACACAGAGTGGAAAGG	356
Qy	135	GGCGACCAACGG---CGCAAGCGGATGCTGCTTCCGATCACTACCGCTCATCATGA	191
Db	357	GAAGTGTCTGGAAATGAGACGCAATGATGATGATGATGATGATGATGATGATGATG	416
Qy	192	GGACTTGTGATCTTGTGACCAAGTACGCGCAAGGCGCTTCTCTTGTGCTGCTGTC	251
Db	417	GGACATCATTTCTGCTGAGGAACTGCACTGCACTGCACTGCTGCTGCTGCTGCTG	476
Qy	252	GGGATCATTTCCCTCTGGGCGGAGGCTGATCCGCTCAAGAGAGGAAATGATGATTA	311
Db	477	CCGGCTCTGGCCACAGGATCCGAGCCGAGCGAGTGAAGAGAGGAAATGATGATTA	536
Qy	312	CAGCAACTGATTTGAGCGCTGTTGAGGCGGCTATCAAGCTTGGCTGATTTGATCA	371
Db	537	CAGTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	596
Qy	372	CTGGGATCTGCTCAGCGGCTTCAAGATGCTATGAGGCTGCTCAAGCTGAGAGGT	431
Db	597	CTGGGATCTGCTCAGCGGCTTCAAGATGCTATGAGGCTGCTCAAGATGAGAGGT	656
Qy	432	CCAGCTGGACTTTGAGCGGATGCGAGTGTGCTTGTGAACGTTTGGGAGCCGATCCA	491
Db	657	CAA---CTACTTCAGAGATGAGCCCAAGCTGATGCTTGAAGGCTTGGGAGCCGATGAA	713
Qy	492	GAACTGATCAGCATTAAGAGGCTTGAAGTATGAGGCTGATGATGATGATGATGATG	551
Db	714	GCACTGATCAGCATTAAGAGTCTGCGGCAATGAGGAAAGGCTTATGAGCGGAGCA	773
Qy	552	CAAGCGCGGAGGAGGAGGAGGATTAAGAGCACTCACAGAGGCAACCTGCGCACTGA	611
Db	774	CAATGCGCGGAGGCTGAGCTCGC-----GGCAGCGG	806
Qy	612	GGCGTGGCTCGCTGGAAGAGCCAGATCATGAGCCATGCGCGCGCTGAGCTTACAG	671
Db	807	CGGTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	866
Qy	672	CAGGACCTTTCGCGCTCGCAAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	731
Db	867	CAGGACCTTTCGCGCTCGCAAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	926
Qy	732	TGAGCGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	791
Db	927	GGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	986
Qy	792	TCACATGCTGCTGCTTGCATTCCTCTT	820

OY 612 GCCGTGCTGCTGGAAAGGCCAGATCATGAGCCATGCCGCGCGCTGCCTACAG 671
DB 807 CCTGTACAGAGCGACACCATCATTTAGGCCACCAAACTGGCATTTATTA 866
OY 672 CAGGACCTTTCGCTCGCAAAAGGCCAGATCGGCATCTCGCTCAACGCGACTACTA 731
DB 867 CACACAGTGGCGAGCAAGAGCTCTGGTGGAAATTTCACTGAACGTGTGCTGGG 926
OY 732 TGAGCCCTGGGACACGACATAGCTCGGACACAGAGGCTGTGAGGACGATGATT 791
DB 927 GGAACCTGTGACATTTGTATACCCCAAGACCTAGAGAGCTGCCGAGATACCTACATT 986
OY 792 TCACATTGCTGTTTCCATCCCATCT 820
DB 987 CTGCTGGGCTGGTTTTCACACCCCATTT 1015

RESULT 14.
ACD23979
ID ACD23979 standard; cDNA; 2016 BP.
XX ACD23979;
AC ACD23979;
XX
XX
XX 26-AUG-2003 (first entry)
DE Novel human secreted and transmembrane protein PRO9820 cDNA.
XX
XX Human; secreted and transmembrane protein; PRO; antiinflammatory;
KM antiarteriosclerotic; cardiact; anti-inferility; anti-HIV; cytosolic;
KM antidiabetic; gene therapy; tumour necrosis factor (TNF)-alpha release;
TNF-alpha release; cell proliferation; cell differentiation;
KM gene expression modulator; proteoglycan release; cytokine release;
KM tumour; inflammatory disease; organ failure; atherosclerosis;
KM cardiac injury; inferility; birth defect; premature aging; AIDS;
KM acquired immunodeficiency syndrome; cancer; diabetic complication;
KM chromosome mapping; gene mapping; pharmaceutical; diagnostic; biosensor;
bioreactor; tissue typing; gene; ss.
XX
OS Homo sapiens.
PN US2003032156-A1.
XX
XX 13-FEB-2003.
PD
XX
XX 06-MAY-2002; 2002US-00140474.
PF
XX
XX 31-MAR-1997; 97WO-US005230.
PR 12-JUN-1998; 98WO-US012456.
PR 14-JUL-1998; 98WO-US014552.
PR 28-AUG-1998; 98WO-US017888.
PR 10-SEP-1998; 98WO-US018824.
PR 14-SEP-1998; 98WO-US019093.
PR 14-SEP-1998; 98WO-US019094.
PR 14-SEP-1998; 98WO-US019177.
PR 15-SEP-1998; 98WO-US019330.
PR 17-SEP-1998; 98WO-US019437.
PR 07-OCT-1998; 98WO-US021141.
PR 29-OCT-1998; 98WO-US022991.
PR 29-OCT-1998; 98WO-US022992.
PR 20-NOV-1998; 98WO-US024855.
PR 01-DEC-1998; 98WO-US025106.
PR 05-JAN-1999; 99WO-US000106.
PR 08-MAR-1999; 99WO-US005028.
PR 10-MAR-1999; 99WO-US005190.
PR 20-APR-1999; 99WO-US008615.
PR 14-MAY-1999; 99WO-US010733.
PR 02-JUN-1999; 99WO-US012252.
PR 01-SEP-1999; 99WO-US020111.
PR 08-SEP-1999; 99WO-US020594.
PR 13-SEP-1999; 99WO-US020944.
PR 15-SEP-1999; 99WO-US021090.
PR 15-SEP-1999; 99WO-US021547.

PR 05-OCT-1999; 99WO-US023089.
PR 29-NOV-1999; 99WO-US028214.
PR 30-NOV-1999; 99WO-US028313.
PR 30-NOV-1999; 99WO-US028409.
PR 01-DEC-1999; 99WO-US028301.
PR 01-DEC-1999; 99WO-US028634.
PR 02-DEC-1999; 99WO-US028551.
PR 02-DEC-1999; 99WO-US028564.
PR 02-DEC-1999; 99WO-US028565.
PR 16-DEC-1999; 99WO-US030095.
PR 20-DEC-1999; 99WO-US030911.
PR 20-DEC-1999; 99WO-US030920.
PR 22-DEC-1999; 99WO-US030970.
PR 30-DEC-1999; 99WO-US031243.
PR 30-DEC-1999; 99WO-US031274.
PR 05-JAN-2000; 2000WO-US000219.
PR 06-JAN-2000; 2000WO-US000277.
PR 06-JAN-2000; 2000WO-US000376.
PR 11-FEB-2000; 2000WO-US003565.
PR 18-FEB-2000; 2000WO-US004341.
PR 18-FEB-2000; 2000WO-US004342.
PR 22-FEB-2000; 2000WO-US004414.
PR 24-FEB-2000; 2000WO-US004914.
PR 24-FEB-2000; 2000WO-US005004.
PR 01-MAR-2000; 2000WO-US005601.
PR 02-MAR-2000; 2000WO-US005746.
PR 10-MAR-2000; 2000WO-US005841.
PR 15-MAR-2000; 2000WO-US005884.
PR 20-MAR-2000; 2000WO-US007377.
PR 21-MAR-2000; 2000WO-US007532.
PR 31-MAR-2000; 2000WO-US008439.
PR 17-MAY-2000; 2000WO-US013705.
PR 22-MAY-2000; 2000WO-US014042.
PR 30-MAY-2000; 2000WO-US014941.
PR 02-JUN-2000; 2000WO-US015264.
PR 28-JUL-2000; 2000WO-US020710.
PR 11-AUG-2000; 2000WO-US022031.
PR 23-AUG-2000; 2000WO-US023522.
PR 24-AUG-2000; 2000WO-US023328.
PR 08-NOV-2000; 2000WO-US030952.
PR 10-NOV-2000; 2000WO-US030873.
PR 01-DEC-2000; 2000WO-US032678.
PR 20-DEC-2000; 2000US-00747259.
PR 20-DEC-2000; 2000WO-US034956.
PR 28-FEB-2001; 2001US-00796498.
PR 28-FEB-2001; 2001WO-US006520.
PR 01-MAR-2001; 2001WO-US006666.
PR 09-MAR-2001; 2001US-00802706.
PR 14-MAR-2001; 2001US-00806889.
PR 22-MAR-2001; 2001US-00816744.
PR 03-APR-2001; 2001US-00828366.
PR 10-MAY-2001; 2001US-00854208.
PR 10-MAY-2001; 2001US-00854280.
PR 18-MAY-2001; 2001US-00860216.
PR 25-MAY-2001; 2001US-00866028.
PR 25-MAY-2001; 2001US-00866034.
PR 25-MAY-2001; 2001WO-US017092.
PR 01-JUN-2001; 2001US-00872035.
PR 01-JUN-2001; 2001WO-US017800.
PR 05-JUN-2001; 2001US-00874503.
PR 14-JUN-2001; 2001US-00882636.
PR 19-JUN-2001; 2001US-00886342.
PR 20-JUN-2001; 2001WO-US019692.
PR 21-JUN-2001; 2001US-00887879.
PR 22-JUN-2001; 2001WO-US020116.
PR 29-JUN-2001; 2001WO-US021066.
PR 09-JUL-2001; 2001WO-US021735.
PR 16-JUL-2001; 2001US-00908827.
PR 06-AUG-2001; 2001US-00924419.
PR 09-AUG-2001; 2001US-00927796.
PR 16-AUG-2001; 2001US-00931836.
PR 19-DEC-2001; 2001US-00028072.

XX (GENTH) GENENTECH INC.
 XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W,
 XX Gerlsten ME, Goddard A, Godowski PJ, Gunney AL, Sherwood S,
 XX Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z,
 XX WPI; 2003-341980/32.
 XX P-PSDB; ABO17742.
 XX
 XX New secreted and transmembrane PRO nucleic acids, for treating
 XX inflammation, organ failure, atherosclerosis, cardiac injury,
 XX infertility, birth defects, premature aging, acquired immunodeficiency
 XX syndrome (AIDS), or cancer.
 XX
 XX Claim 2; Fig 253; 660bp; English.
 XX
 XX The invention describes an isolated nucleic acid (1) comprising, or which
 XX has 80 % sequence identity to, or the full-length coding sequence of, one
 XX of 275 nucleotide sequences, and which encodes a corresponding
 XX polypeptide selected from 275 amino acid sequences, where all sequences
 XX are given in the specification. The polypeptide encoded by (1) is used to
 XX detect PRO polypeptides, link a bioactive molecule to a cell expressing a
 XX PRO polypeptide, modulate a biological activity of a cell, stimulate the
 XX release of tumour necrosis factor (TNF)-alpha from human blood, modulate
 XX the uptake of glucose or free fatty acid by cells, stimulate or inhibit
 XX the proliferation or differentiation of cells or gene expression,
 XX stimulate the release of proteoglycans, inhibit the binding of cytokine
 XX to factor VIIa, or detect the presence of tumour in a mammal. The nucleic
 XX acid and polypeptide encoded by 1c, are useful for treating inflammatory
 XX diseases, organ failure, atherosclerosis, cardiac injury, infertility,
 XX birth defects, premature aging, acquired immunodeficiency syndrome
 XX (AIDS), cancer, or diabetic complications. The nucleic acid is useful as
 XX hybridisation probes, in chromosome and gene mapping, and in generating
 XX antisense RNA or DNA. The polypeptides are useful as pharmaceuticals,
 XX diagnostics, biosensors or bioreactors. Both are useful in tissue typing.
 XX This sequence encodes a novel human secreted and transmembrane PRO
 XX polypeptide
 XX
 XX Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;
 XX
 XX Query Match 12.2%; Score 178.2; DB 7; Length 2016;
 XX Best Local Similarity 54.5%; Pred. No. 1.5e-32;
 XX Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;
 XX
 XX 15 AGCTTGGCCCAACGATTGATGGGCTTGCAACGCGCCCTTACCGATCGAAGGCGC 74
 XX 237 AACCTTCCCTCTGCTTCTCTGAGGCTGGGCACTTCTGCTACGACGAGGCGC 296
 XX 75 CGTCAAGAGGTGGCCGCGCCCTTCATCTGGGACAGTATGCGACCTGGAGCCATC 134
 XX 297 CTGGGACCAAGGCGGGAAGGGCTTACGATCTGGGACGCTTCAACACAGTGGGAGG 356
 XX 135 GCGCACCAAGG---CGCAACGCGGATGGCTTGGCATCACTACCCGCTACGATGA 191
 XX 357 GAAAGTGTCTGGGAATGAGAGCGGAGATGAGCCGTGACGCGTACTCAAGGTCACGA 416
 XX 192 GAACCTTGTATCTCTTGAACCAAGTACGCGCAAGGCTTACCGCTTCTGTGTGTC 251
 XX 417 GGACATCATCTTCTGCTGAGGAGAACTGACGTACCACTACCGATTTCTCTGTGGCC 476
 XX 252 GCGGATCATTCCTCCCTCGGCGGCAAGGCTGATCCGCTCAACGAGAGGGAATTGATT 311
 XX 477 CCGGCTCTGCGCCCAAGGCAATCCGAGCCGAGCAGTGAACAAAGGGAATCGAATCTA 536
 XX 312 CAGCAAACTGATTAACCGCCCTTGTGAGCGGGGATATCAACGCTTGGGATTTGTACCA 371
 XX 537 CAGGATCTTATCGATGCTCTTCTGAGCAAGCAATCACTCCATCGGACCTTGCACCA 596
 XX 372 CTGGGATCTGCTCAGGCGCTTACGATCGCATGAGAGGCTGCAACGTAAGAGGT 431
 XX 597 CTGGGATCTGCAACAGCTGCTCCAGGTCAATACGCTGGTGGGAGAAATGTAGCATGGC 656

QY 432 CCAGCTGACCTTTGAGCGGATATGAGAGTGTGCTTTGAACGTTTGGGACGAGTCCA 491
 DB 657 CAA---CTACTTCAAGAACTACACGCAACCTGTGCTTTGAGCCCTTGGGACCGTGTGAA 713
 QY 492 GAACGTGATCAACATCAACGAGCCCTGATTCAGGCCATGTATGATATGCCACCGGAG 551
 DB 714 GCATCGATCAACGTTTCAAGTATCTCTCGGCAATGGCGAATAAAGGCTATGAGACGGGCCA 773
 QY 552 CAAGCCCCGGGCGAGGAGCAATTAACAAGCACTCCACGAGGGCAACCTGCCACTGA 611
 DB 774 CCAATGCGCGGCGCTGAAGCTCCGC-----GGCACGG 806
 QY 612 GCCGTGCTGCTGGAAGGCGCAAGATCATGAGCCATGCCGCGCTGCGCTTACAG 671
 DB 807 CTGTCAAGAGGAGCAACCAACATCTTAAGGCCACGCAAAACCTGGATCTTATTA 866
 QY 672 CAGGACCTTTCGCCCCCTGCAAAAAGGCGCATGCGATCTGCTCAACGCGACTATCTA 721
 DB 867 CACCACGTGCGCGAGCAAGGAGCAAGGCTGTGGGAAATTTCACTGAATGTGAGG 926
 QY 732 TGAGCCCTGGGACAGCAATAGCCTGGGACAGAGAGGCTGTGAGAGCATGGAATT 791
 DB 927 GGAACCTGTGACATTAATGTAACCCCAAGGACTTGAAGGCTGCCAGATTACTTACAGTT 986
 QY 792 TCACATTTGGCTGTTTGGCCAATCCATCT 820
 DB 987 CTGTCTGGGCTGTGTTTCCCAACCCATTT 1015
 XX
 XX RESULT 15
 XX ACA67120
 XX ID ACA67120 standard; cDNA; 2016 BP.
 XX AC
 XX ACA67120;
 XX DT 23-JUN-2003 (first entry)
 XX XX
 XX cDNA encoding human PRO polypeptide #127.
 XX
 XX Human: PRO polypeptide; secreted and transmembrane protein;
 XX anti-PRO antibody; diagnostic assay; gene expression; diabetes;
 XX bone disorder; cartilage disorder; rheumatoid arthritis; Obesity;
 XX sports injury; osteoarthritis; hyper-insulinaemia; hypo-insulinaemia;
 XX hearing loss; coagulation disorder; stroke; heart attack; cardiac;
 XX anti-diabetic; anorectic; vulnery; antiarthritic; osteopathic;
 XX antineumatic; auditory; cerebroprotective; angiogenic; gene; ss.
 XX
 XX Homo sapiens.
 XX OS
 XX US200304311-A1.
 XX PN
 XX PD 02-JAN-2003.
 XX
 XX PF 19-DEC-2001; 2001US-00028072.
 XX XX
 XX 18-JUN-1997; 97US-0049911P.
 XX 26-AUG-1997; 97US-0056974P.
 XX 17-SEP-1997; 97US-0059113P.
 XX 17-SEP-1997; 97US-0059115P.
 XX 17-SEP-1997; 97US-0059117P.
 XX 17-SEP-1997; 97US-0059122P.
 XX 17-SEP-1997; 97US-0059184P.
 XX 18-SEP-1997; 97US-0059163P.
 XX 18-SEP-1997; 97US-0059352P.
 XX 19-SEP-1997; 97US-0059588P.
 XX 24-SEP-1997; 97US-0059636P.
 XX 17-OCT-1997; 97US-0062250P.
 XX 17-OCT-1997; 97US-0062287P.
 XX 17-OCT-1997; 97US-0062287P.
 XX 17-OCT-1997; 97US-0063755P.
 XX 24-OCT-1997; 97US-0062814P.
 XX 24-OCT-1997; 97US-0062816P.

PR 24-OCT-1997; 97US-0063045P.
 PR 24-OCT-1997; 97US-0063082P.
 PR 24-OCT-1997; 97US-0063127P.
 PR 27-OCT-1997; 97US-0063327P.
 PR 27-OCT-1997; 97US-0063329P.
 PR 28-OCT-1997; 97US-006350P.
 PR 28-OCT-1997; 97US-0063561P.
 PR 29-OCT-1997; 97US-0063704P.
 PR 29-OCT-1997; 97US-0063733P.
 PR 29-OCT-1997; 97US-0063735P.
 PR 29-OCT-1997; 97US-0063738P.
 PR 03-NOV-1997; 97US-0064248P.
 PR 07-NOV-1997; 97US-0064809P.
 PR 12-NOV-1997; 97US-0065186P.
 PR 17-NOV-1997; 97US-0065466P.
 PR 21-NOV-1997; 97US-0066364P.
 PR 24-NOV-1997; 97US-0066453P.
 PR 24-NOV-1997; 97US-0066511P.
 PR 24-NOV-1997; 97US-006670P.
 PR 11-DEC-1997; 97US-0069212P.
 PR 11-DEC-1997; 97US-0069278P.
 PR 11-DEC-1997; 97US-0069334P.
 PR 16-DEC-1997; 97US-0069694P.
 PR 23-JAN-1998; 98US-0072320P.
 PR 04-FEB-1998; 98US-0073612P.
 PR 09-FEB-1998; 98US-0074082P.
 PR 12-MAR-1998; 98US-0077791P.
 PR 20-MAR-1998; 98US-0078910P.
 PR 25-MAR-1998; 98US-0079294P.
 PR 27-MAR-1998; 98US-0079663P.
 PR 27-MAR-1998; 98US-0079728P.
 PR 31-MAR-1998; 98US-0080165P.
 PR 12-JUN-1998; 98MO-US012456.
 PR 14-JUL-1998; 98MO-US014552.
 PR 28-AUG-1998; 98MO-US01788P.
 PR 10-SEP-1998; 98MO-US018824.
 PR 14-SEP-1998; 98MO-US019093.
 PR 14-SEP-1998; 98MO-US019094.
 PR 14-SEP-1998; 98MO-US019177.
 PR 16-SEP-1998; 98MO-US019330.
 PR 17-SEP-1998; 98MO-US019437.
 PR 07-OCT-1998; 98MO-US021141.
 PR 29-OCT-1998; 98MO-US022891.
 PR 29-OCT-1998; 98MO-US022892.
 PR 20-NOV-1998; 98MO-US024855.
 PR 01-DEC-1998; 98MO-US025108.
 PR 05-JAN-1999; 99MO-US000106.
 PR 08-MAR-1999; 99MO-US005028.
 PR 10-MAR-1999; 99MO-US005190.
 PR 20-APR-1999; 99MO-US008615.
 PR 14-MAY-1999; 99MO-US010733.
 PR 02-JUN-1999; 99MO-US012252.
 PR 01-SEP-1999; 99MO-US020111.
 PR 08-SEP-1999; 99MO-US020594.
 PR 13-SEP-1999; 99MO-US020944.
 PR 15-SEP-1999; 99MO-US021090.
 PR 15-SEP-1999; 99MO-US021547.
 PR 05-OCT-1999; 99MO-US023089.
 PR 29-NOV-1999; 99MO-US028214.
 PR 30-NOV-1999; 99MO-US028313.
 PR 30-NOV-1999; 99MO-US028409.
 PR 01-DEC-1999; 99MO-US028301.
 PR 01-DEC-1999; 99MO-US028634.
 PR 02-DEC-1999; 99MO-US028551.
 PR 02-DEC-1999; 99MO-US028564.
 PR 16-DEC-1999; 99MO-US028565.
 PR 20-DEC-1999; 99MO-US030911.
 PR 20-DEC-1999; 99MO-US030939.
 PR 30-DEC-1999; 99MO-US031243.
 PR 30-DEC-1999; 99MO-US031274.
 PR 05-JAN-2000; 2000MO-US000219.

PR 06-JAN-2000; 2000MO-US000277.
 PR 06-JAN-2000; 2000MO-US000376.
 PR 11-FEB-2000; 2000MO-US003565.
 PR 18-FEB-2000; 2000MO-US004341.
 PR 18-FEB-2000; 2000MO-US004342.
 PR 22-FEB-2000; 2000MO-US004414.
 PR 24-FEB-2000; 2000MO-US004914.
 PR 24-FEB-2000; 2000MO-US005004.
 PR 01-MAR-2000; 2000MO-US005601.
 PR 02-MAR-2000; 2000MO-US005746.

(GETH) GENENTECH INC.

PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gertlisen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

DR WPI; 2003-352836/33.
 DR P-PSDB; ABUS0996.

PT New isolated PRO polypeptide useful for treating diabetes, rheumatoid
 PT arthritis, sports injuries, obesity, hearing loss in mammals, stroke, or
 PT heart attack.

Claim 2; Fig 253; 643pp; English.

CC The present invention relates to the isolation of novel human PRO
 CC polypeptides, and the polynucleotide sequences encoding them. The PRO
 CC polypeptides are secreted and transmembrane proteins. The PRO
 CC polypeptides and polynucleotides are useful for preparing a medicament
 CC useful in the treatment of diabetes, bone and/or cartilage disorders
 CC (e.g. rheumatoid arthritis, sports injuries, osteoarthritis), obesity,
 CC hyper- or hypo-insulinaemia, hearing loss, and coagulation disorders
 CC (e.g. stroke, heart attack). Anti-PRO antibodies are useful in diagnostic
 CC assays for PRO, by detecting its expression in specific cells, tissues or
 CC serum, and for affinity purification of PRO from recombinant cell culture
 CC or natural sources. ACA65994-ACA67268 represent cDNA sequences encoding
 CC the human PRO polypeptides of the invention. Note: The sequence data for
 CC this patent was obtained in electronic format directly from the USPTO web
 CC site at seqdata.uspto.gov/patseq/identity.html
 CC
 XX

Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

Query Match 12.2%; Score 178.2; DB 7; Length 2016;
 Best Local Similarity 54.5%; Pred. No. 1.5e-32;
 Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

QY 15 AGCTTGCCCAAGCACTTGAATGGGCTTGCAGACGGCCCTTACAGATCGAAGCCG 74
 DB 237 AACCTTCCCTCTGGGCTTCTCGGGGCGTGGGACATTTGCTTACAGACGAGGCGC 286
 QY 75 CGTCAAGAAAGGTGGCGCGCCGCTTCATCTGGGACAGCTACTGCACTTGAGCCATC 134
 DB 297 CTGGGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 356
 QY 135 GCGCACAACCG--GCGCAAGCGATGGTGGCTTGCATCTACACAGCTTCAAGTGA 191
 DB 357 GAAAGTCTTGGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 416
 QY 192 GGAATTTGATCTTGTACCAAGTACGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 251
 DB 417 GGAATTTGATCTTGTACCAAGTACGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 476
 QY 252 GCGGATCTTCCCTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 311
 DB 477 CCGGCTCTGCGCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 536
 QY 312 CAGCAACTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 371
 DB 537 CAGTGAATTTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 596
 QY 372 CTGGGATCTGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 431

```
Db 597 CTGGATCTGCCACAGCTGCTCCAGTCAAAATACGGTGGGTCAGAAATGTAGCATGAC 656
QY 432 CCAGCTGGACTTTGAGCCGTTATGAGGTTGTGCTTTGAAAGTTTGGGGAACCGAGTCCA 491
Db 657 CAA---CTACTTCAGAGACTAGCCACCTGTGCTTTAGGCTTTGGGGACCGTGTGA 713
QY 492 GAACTGGATCAACCATCAAGAACCCCTGGATTTCAGGCCATCTATGGATATGCCACCGGAG 551
Db 714 GCACGTGATCAAGTTCAAGTATCTCGGGCAATGCGAAGAAAGCTATGAGACGGGCCA 773
QY 552 CAACGCCCCGGGCGAGGACGATTAACAAGCACTCCACCGAGGGCAACACTGCCACTGA 611
Db 774 CCATCGCGCGGGCTGAAGCTCCGC-----GGCACCGG 806
QY 612 GCCGTGGCTCGCTGGAAGGCCCCAGATCATGAGCCATGCCCGCCGTGGCCGTCTACAG 671
Db 807 CCTGTACAAGGCGAGCAACACATCATTTAAGGCCCAACCAAACTTGGCATTTCTTATTA 866
QY 672 CAGGACTTTGCCCCCTGCAAAAGGGCCAGATCGGCATCTCGCTCAACGGCGACTACTA 731
Db 867 CACCACTGGCGCAGCAAGCAAGGTCTGGTGGAAATTTCACTGAACGTGACTGGGG 926
QY 732 TGAAGCCCTGGGACAGCATGAGCCCTCGGGAACAAGAGGCTGTGAGCGACGATGAAATT 791
Db 927 GGAACCTGTGACATTAAGTAACCCCAAGAGACTTAAGGCTGCCGAGAGATACCTACAGTT 986
QY 792 TCACATTGGCTGTTTGGCCAATCCCATCT 820
Db 987 CTGTCTGGGCTGTGTTTGCCAACCCCATTT 1015
```

Search completed: March 28, 2004, 17:03:18
Job time : 587.342 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 20:52:22 ; Search time 105.557 Seconds

(without alignments)
7649.443 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455
Sequence: 1 atgccgcagtcgctagctctc.....gggttaagtgcgcgcataa 1455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	10.3	1521	4	US-09-914-841A-14
2	139.4	9.6	1467	4	US-09-489-039A-2386
3	130.8	9.0	1317	4	US-09-134-078-2
4	128.2	8.8	3163	4	US-09-344-510B-6
5	128.2	8.8	3435	4	US-09-344-510B-7
6	127.8	8.8	1630	4	US-09-344-510B-9
7	127.8	8.8	5032	4	US-09-344-510B-8
8	126.2	8.7	1287	4	US-09-914-841A-2
9	124.8	8.6	1542	2	US-09-122-230-8
10	124.8	8.6	1909	2	US-09-122-230-6
11	118.6	8.2	1829	3	US-08-688-988-9
12	111.4	7.7	1545	1	US-08-400-275-17
13	104.8	7.2	1458	4	US-09-134-000C-2890
14	101.4	7.0	1437	4	US-09-107-532A-1268
15	95.6	6.6	1410	4	US-09-489-039A-1053
16	93.8	6.4	1407	3	US-08-688-988-7
17	90.4	6.2	2435	3	US-09-306-593-1
18	88.4	6.1	1931	4	US-09-431-470-1
19	88.4	6.1	1931	4	US-09-431-470-3
20	87.6	6.0	454	4	US-09-615-192A-166
21	87.6	6.0	454	4	US-09-169-789-166
22	85.2	5.9	457	2	US-08-975-316-79
23	85.2	5.9	457	2	US-09-615-192A-79
24	85.2	5.9	457	4	US-09-169-789-79
25	85.2	5.9	470	4	US-09-615-192A-141
26	85.2	5.9	470	4	US-09-169-789-141
27	84.8	5.8	3460	4	US-09-344-510B-10

C	28	84.2	5.8	1404	4	US-09-489-039A-4718	Sequence 4718, Ap
	29	84.2	5.8	1455	4	US-09-489-039A-4800	Sequence 4800, Ap
	30	79.4	5.5	1443	4	US-09-489-039A-1986	Sequence 1986, Ap
C	31	74.8	5.1	16535	4	US-08-961-527-74	Sequence 74, Appl
	32	71	4.9	1428	4	US-09-489-039A-2983	Sequence 2983, Ap
	33	69.4	4.8	1479	4	US-09-134-000C-289	Sequence 289, Appl
	34	68.4	4.7	286	4	US-09-313-294A-7321	Sequence 7321, Ap
C	35	68	4.7	3615	1	US-08-920-812-17	Sequence 17, Appl
	36	68	4.7	3615	1	US-08-920-827-17	Sequence 17, Appl
C	37	68	4.7	3615	1	US-08-921-177-17	Sequence 17, Appl
	38	68	4.7	3615	1	US-08-362-577C-17	Sequence 17, Appl
C	39	68	4.7	3615	2	US-08-961-527-115	Sequence 115, App
C	40	63	4.3	11303	4	US-08-961-527-115	Sequence 115, App
	41	57.6	4.0	1371	4	US-09-134-000C-1598	Sequence 1998, Ap
	42	56.4	3.9	1464	4	US-09-489-039A-4740	Sequence 4740, Ap
	43	54.6	3.8	1404	4	US-09-489-039A-5778	Sequence 5778, Ap
	44	54.2	3.7	1455	4	US-09-489-039A-2724	Sequence 2724, Ap
	45	54	3.7	1424	4	US-09-615-192A-138	Sequence 138, App

ALIGNMENTS

```
RESULT 1
US-09-914-841A-14
; Sequence 14, Application US/09914841A
; Patent No. 6645750
; GENERAL INFORMATION:
; APPLICANT: Amano Enzyme Inc.
; TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE
; FILE REFERENCE: 066072
; CURRENT APPLICATION NUMBER: US/09/914,841A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: P. Hei. 11-056299
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: PCT/JP00/01242
; PRIOR FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: Camellia var. sinensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1521)
; OTHER INFORMATION:
; NAME/KEY: mat.peptide
; LOCATION: (235)..()
; OTHER INFORMATION:
US-09-914-841A-14

Query Match      10.3%; Score 150; DB 4; Length 1521;
Best Local Similarity 52.2%; Pred. No. 3.3e-29;
Matches 455; Conservative 1; Mismatches 401; Indels 15; Gaps 5;

QY      20  TGGCCACGACCTTTGAATGGGCGCTTGGCAAGCGCGCTTACGATGGAAGGCGCGTCA 79
DB      116  TTCTGATGTTTGTCTTGGAGCTGCTCTTCTGCTTACGATGGAAGGCGCGTCA 175

QY      80  AAGAGGTGGCGCGCGCGCTGCATCGGAGACGTAATGCCACCGGAGCGATCGGCA 139
DB      176  AGAGAGGTGGAGAAAGGCCCAATTTTGGATTCCTTCACTCATGAGTTTCCAGGTA 235

QY      140  ---CCAAGCGCGCAACGCGATGTGGCTTGCATCACTACCAACGCTACGATGAGACT 196
DB      236  TATCGATGTGTACACGAGATGTAGCTGATACCTTTATCACTGTTACAAAGGATG 295

QY      197  TTGATCTCTTGACCAAGTACGCGCAAGGCGCTTCTCTTGTGTGTGGCGGA 256
DB      296  TGAAGGTGCTGGAAGTTTATGAGACTAGATGTTTCAGAAATGTCATCTCATGCGGCGAG 355

QY      257  TCATTTCCCTCGGCGGAGGCTGATCCGCTCAACGAGGAGGAATGATTTACAGCA 316
```

Db 356 TATTACCTGGGGAGAGCTTAGACGAGAGAGTGAACAAGAGATATGCTTTCACACA 415
Qy 317 AACTGATTTAGACCTCTGTTAGAGGGGATATACAGCCTTGGGTGACTTTGATACACTGG 376
Db 416 ATGCATCAATAGACCTTTATTCGAAGGTATACCACTTTTATTAACATCTTTCACTGGG 475
Qy 377 ATTCGCTCAGGCGCTTACAGATGCGCTATGAGAGGTGGCTCAAGTGGAGAAGGCCAGC 436
Db 476 ATCTTCCCAAGCCCTAGAGATGATATGAGAGCTTTTAAAGC---CCACACATTGGA 532
Qy 437 TGAAGCTTGAAGCGGTATGCGAGTTGTGCTTTGAACGTTTGGGAGCCGAGTCCAGACT 496
Db 533 ACGATTTCCGGGATTTTGCACAGCGTGTCTTCAAGAGATTGGTGAACGAGTTTAAAGATT 592
Qy 497 GGATCAACATCAAGAGACCTTGAGTTCAAGGCATCTATGATATGCCAGCGACAGCAAG 556
Db 593 GGATCAACATCAAGAGACCTATGCTTACTCTATGGGGTTATGATGACAGTCTCTAG 652
Qy 557 CCCCCGGCAGAGAGCATTAACAAGCACTC---CACGAGGGGCAACATGCACTGAGC 613
Db 653 CACCGGCGCTTTGCTTGGCTTTTATGSCATTTTGCCTTAAAGGAGATTCTGGAGCTGAGC 712
Qy 614 CGTGGCTGCTGGAAGGCGCCAGATCATGAGCCATGCCGCGCGCTGAGCCGCTTACAGCA 673
Db 713 CCTATATGATTAACCAACAATTTGCTTTCTCATGCTGCTGCTGTAACATATACAGG 772
Qy 674 GGGACTTTCGCCCTCGCAAAAGGCGCGATCGGATCTGCTCAAGCGGCACTATG 733
Db 773 AGAATATACAGGCATATCAAAAGGGGAGATAGGATTAACACTGATGACTTATGATGA 832
Qy 734 AGCCTGGGACAGCAATGAGCCTCGGACAGAGAGCTGTGAGGAGCGATGGAATTC 793
Db 833 TTCCCT---ACTCCAATTCGAAAGCCGACAGAGATGACGACACAGACCCTTGAATTCA 889
Qy 794 ACATTGCTGTTGGCCATTCATCTTTGAAGAGAGACTATCCAGAGCATGAAGA 853
Db 890 TGTATGATGTTTATGAGCATTAAGCTT---TGTGATATTCACAAAGACATGCGTA 946
Qy 854 AGCAGCTGGGAGAGGCTTCCAGCCCTCACT 885
Db 947 GACTCGTTGTAAAGGTTACCAAGTTCACT 978

RESULT 2
US-09-489-039A-2386
; Sequence 2386, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2386
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2386

Query Match 9.6%; Score 139.4; DB 4; Length 1467;
Best Local Similarity 56.1%; Pred. No. 1.8e-26;
Matches 307; Conservative 1; Mismatches 227; Indels 12; Gaps 2;

Qy 16 GCTCGCCCAAGCATTTGAATGGGGCTTGCACAGCGCGCTACAGATGAAGGGCGCC 75
Db 103 GCTTTCCGCAATTTCTTATGGGGTGCCTTACGCGATCATGAGTTGAGGAGGA 162
Qy 76 GTCAAGAAGGTGGCGGCGCGCTCATCTGGGACAGTACTGCACTGAGGCATCG 135

Db 163 CAGAGCCGACCGTAAAGAGACCGTCAATCTGGGATATCTACTCTCATCTGCCGGTACC 222
Qy 136 CGCACACAGCGCCCAACCGCGATGTGGCTTGGATATCAACACCGCTACGATGAGAC 195
Db 223 ACCTTTAAAGGACACACCGCGATATTTGCCGTGACACATATACACGTTTCCGGAAGC 282
Qy 196 TTTGATCTTGAACCAATGACGGGCGCAAGAGCCCTACCGCTTCTCTGTGTGGTCCGG 255
Db 283 GTGGCGTTAATGGCCGAGATGGGCGTGCACAAAGCTATGATTCATCTCTGGCCCCG 342
Qy 256 ATCATTCCTCGCGCGAGCTGAGATCCGCTCAACGAGAGAGGAATTAAGTTTACAGC 315
Db 343 CTGCTGCC-----CGCGGGCGGGTAAAGTGAATGAACGGGAGTCCAGTTTACAGC 396
Qy 316 AAACATGATACCGCCCTGTTGAGCGGGGATATACCGCTTGGGTGATTTGTACACTGG 375
Db 397 GATCTGATGACAGAACTGTGGCGCACAAATATGAGCCGATGATCACTTATCACTAG 456
Qy 376 GATCTGCTCAGGCGCTTCAAGATCGCTATGAGAGCTGCTCAACGTGAAGAGTCCAG 435
Db 457 GATCTGCCGAGCGCCCTGCAAG-----ATGAGGGCGGCTGGGAAGGCCACATACCGCC 510
Qy 436 CTGACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGACCGAGTCCAGAAC 495
Db 511 GAGGCTTCCCGAGTACGCGCCGCTGTATGCGCGCTTTGGTTCAAGGGTGAACCTG 570
Qy 496 TGTATCAACATCAAGAGCCCTGATTCAGGCCATCTATGATATGCAACCGGAGAAC 555
Db 571 TGGGCTACTTCAACGAAACCATTTGTTCAATTGGGACGGCTACATTAAGGCTTCAT 630
Qy 556 GCCCGCG 562
Db 631 CCCCCG 637

RESULT 3
US-09-134-078-2
; Sequence 2, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halie, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 1317 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...1314
US-09-134-078-2

Query Match 9.0%; Score 130.8; DB 4; Length 1317;
Best Local Similarity 56.4%; Pred. No. 2.8e-24;
Matches 309; Conservative 1; Mismatches 223; Indels 15; Gaps 3;

16 GCCTGCGCCACGACCTTTGAATGGGCTTGCGCAAGCGCCCTTACCAAGATCCAAAGCGCC 75
16 GATTTTCCAAAAGATTTTATCTTCGGAACGGCTACCGACATACAGATTGAAGTGCA 75
76 GTCAAGAAAGGTGGCGCGCGCCGTCATCTGGGACACGACTGCGCACTGAGCCATCG 135
76 GCAAAGAGAGATGGCAGAGGCGCATCATTTGGGATGCTTTTACACAGCGCTGGCA 135
136 CGCAGCAACGCGCCCAACGCGCATGTGGCTTGATCACTACCAACGCTACGATGAGAC 195
136 ACCCTGACGGTGACACAGAGAGCGTTGCGTGACATATACCGATACAGGAAGAT 195
196 TTGATCTCTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTTGTGTGTGCGCG 255
196 ATCCACTGATGAAAGAAATAGGTTAGACGCTTACAGGTTCTTCTATCTCTGCGCCAGA 255
256 ATCATTCCTCCGCGCGGCGAGCGTGCATCCGCTCAACGAGAGGAAATGAGTTTACAGC 315
256 ATTATGCCAGATG-----GAAGACATACCAACAAAGGTTGATTTCTACAC 306
316 AAACGATGACGCGCTGTTAGCGGCGGTATCACCGCTTGAGTGTGATTTGACACTGG 375
307 AGACTGCTGATGAGCTTTGAAAGATGATATCATACATTCGTAACACTCTATCACTGG 366
376 GATCTGCTGAGCGCTTACGATCGCTATGAGGCTGCTCAACGTCGAGAGGTCAG 435
367 GACTTACCTTACGACCTTATGAAA--AAGTGAGTGGCTTAAACC--AGATATAGCG 420
436 CTGAGCTTTGAGGGTATGCGAGTGTGTGCTTTGAAGCTTTGGGACGAGTCCGAAC 495
421 CTCTATTACAGACATACGACGCTTATGTCAACGACCTCGGTATCTGTGAACAT 480
496 TGATCACCATCAACGAMCCCTGATTCAAGCCATCTATGATATGCCACGCGACGAC 555
481 TGATTAACACTGAACGACATGTGTCTTTCTCGGTTATTAACAGGAGAGCAT 540
556 GCCCGCGG 563
541 GCCCGCGG 548

RESULT 4
US-09-344-510B-6
Sequence 6, Application US/09344510B
Patent No. 6579850
GENERAL INFORMATION:
APPLICANT: Nabeshima, Youichi
Kuroo, Makoto
Sekine, Susumu
Iida, Akihiro

TITLE OF INVENTION: No. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States

ZIP: 10112-3801
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 1.44 mb, DS, DD
COMPUTER: Compaq DeskPro EN
OPERATING SYSTEM: Windows 98
SOFTWARE: Wordpad
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/344, 510B
FILING DATE: 25-Jun-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/04585
FILING DATE: 12-DEC-1997
APPLICATION NUMBER: JP 347871
FILING DATE: 26-DEC-1996
APPLICATION NUMBER: JP 205815
FILING DATE: 31-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Peiray, Lawrence S.
REGISTRATION NUMBER: 31865
REFERENCE/DOCKET NUMBER: 766.32
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 218-2100
TELEFAX: (212) 218-2200
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3163
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: human
IMMEDIATE SOURCE:
LIBRARY: kidney
FEATURE:
NAME/KEY: CDS
LOCATION: 9...3047
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-344-510B-6

Query Match 8.8%; Score 128.2; DB 4; Length 3163;
Best Local Similarity 52.1%; Pred. No. 1.8e-23;
Matches 407; Conservative 0; Mismatches 338; Indels 36; Gaps 4;

130 CCATCGGCGACCAACGCGCGCAACGCGCATGTGGCTTGCGATCACTACCGCTACGAT 189
348 CCGTCCCGCTGACGCGCGCACCGGAGAGTACGACGACGACCAACGCTTTC 407
190 GAGGACTTGTATCTTGAACCAAGTACGGCGCAAGAGCCCTAACCGCTTCTCTGTGTGG 249
408 CGCGACACGAGGCGCTGCGAGCTCGGGGTATCATCACTACCGCTTCTCATCTGTGG 467
250 TCGGATCATTTCCCTCGCGCGCAGCTGATCCGTCACAGAGAGGGAATTGATTT 309
468 GCGGAGTGTCTCCCAATGACAGCGGCGGTGCC--CAACGCGAGGGGCTGCGCTAC 524
310 TACAGCAACTGATTGACGCGCTTGTGAGCGGGGTATCACGCTTGGGTGATCTTTGAC 369
525 TACCGCGCTGCTGAGCGGCTGCGGAGCTGGGCGGTGACGCGCTGTGACCTGTAC 584
370 CACTGGATCTGCTCAGAGCGCTTCAAGATCGCTATGAGGCTGGGCTCAACGTCGAAGAG 429
585 CACTGGAGCTGCGCCAGCGCTGACAGAGAGCTTACGCGGCTGGGCAACCGCGCTTG 644
430 GTCCAGCTGACTTGTAGCGGATGACAGGTTGTGCTTGAACGTTTGGGGAACGAGTC 489
645 GCGGAC--ACTTACAGGATTAAGCGGAGCTGTGCTTCCGCACTTCCGCGGTACGTC 701
490 CAGACTGATCACTCAACGAMCCCTGATTCAGGCCATCTATGATATGACACCGGC 549
702 AAGTACTGATCACTCAACGAMCCCTTACGTGTGTGCTGCGGACGCTACCGCACCGGG 761

QY 550 AGCAAGCCCGGCGAGAGCAGATTAAACAGCACTTCAACGAGGCAACACTGCCACT 609
 Db 762 CGCTGGCCCCCGGCAATCCGGGGGAGCCGCGCTC----- 797
 QY 610 GAGCGGTGCTCGCTGGAAAGGCCAGATCATGAGCCATGCCGCCGCGCTGCTAC 669
 Db 798 ---GGTACCTGGGGGAGCAACCTCTCTGCTGCTGCAAGGCTGGCATCTCTAC 854
 QY 670 AGCAGGAGCTTTGGCCCCGCGCAAAAGGCCAGATCGGCTCAACGAGGAGTAC 729
 Db 855 AATCTCTTCTTCCTCCACTCAGGAGGTGAGGTGCTCATGCTTCAAGCTTCACCTGG 914
 QY 730 TATGAGCCTTGGGACAGCAATGAGCTTCGGGCAAGAGGCTGTCGACGAGTGA 789
 Db 915 ATCAATCTCGAAGAAATGACCGACCAAGCATCAAAAGATGCAAAATCTC---TGGAC 971
 QY 790 TTTCACTTGGCTGTGTTGGCAATCCCATCTTCTTGAAGAAGCTTCCAGAGAGCATG 849
 Db 972 TTTGTACTAGGTGTGTTGGCAACCCGTATTTATGATGCTATCCCGAGAGCATG 1031
 QY 850 AAGAAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGCCATCTCAAT 909
 Db 1032 AAGATTAACCTTTCACTATTTCTGCTGATTTTACTGATTTGAAGAAAGTTATCAAAA 1091
 QY 910 G 910
 Db 1092 G 1092

RESULT 5
 US-09-344-510B-7
 : Sequence 7, Application US/09344510B
 : Patent No. 6579850

GENERAL INFORMATION:

APPLICANT: Nabeshima, Youichi

Kuroo, Makoto

Sekine, Susumu

Iida, Akihiro

TITLE OF INVENTION: No. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: United States

ZIP: 10112-3801

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 1.44 mb, DS, DD

COMPUTER: Compaq DeskPro EN

OPERATING SYSTEM: Windows 98

SOFTWARE: Wordpad

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/344,510B

FILING DATE: 25-Jun-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP97/04585

FILING DATE: 12-DEC-1997

APPLICATION NUMBER: JP 347871

FILING DATE: 26-DEC-1996

APPLICATION NUMBER: JP 205815

FILING DATE: 31-JUL-1997

ATTORNEY/AGENT INFORMATION:

NAME: Peiry, Lawrence S.

REGISTRATION NUMBER: 31865

REFERENCE/DOCKET NUMBER: 766.32

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 218-2100

TELEFAX: (212) 218-2200

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 3435

TYPE: nucleic acid

STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: human
 IMMEDIATE SOURCE:
 LIBRARY: kidney
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 9..1655
 IDENTIFICATION METHOD: E
 SEQUENCE DESCRIPTION: SEQ ID NO: 7:
 US-09-344-510B-7

Query Match 8.8%; Score 128.2; DB 4; Length 3435;
 Best Local Similarity 52.1%; Pred. No. 1.8e-23;
 Matches 407; Conservative 0; Mismatches 338; Indels 36; Gaps 4;

QY 130 CCATCGGCACCAACGCGCCCAACGCGATGTGGCTTGCATCAACACCGCTACAT 189
 Db 348 CCGTCGCGCTGACAGCCCGCCACCGGGAGCTAGCCAGCAGCTACAAACAGTCTTC 407
 QY 190 GAGGACTTTGATCTTTGACCAAGTACGCGCAGAAAGCTTACCGCTTCTGTGCTGG 249
 Db 408 CGGACACGAGAGCGCTGCGCGAGCTCGGGGTCACTACACGCTTCTCATCTGTGG 467
 QY 250 TCGGAGATCAATCCCTCGCGGCGAGGCTGATCCGTCACGAGAGGGAATGAGTTT 309
 Db 468 GCGGAGTGTCTCCCAATGAGCAGCGGGCTCC---CAACGCGAGGGGCTGCGCTAC 524
 QY 310 TACAGCAACTGATTTGACGCGCTTTGAGGCGGGGTATCAAGGCTTGGTGACTTTTAC 369
 Db 525 TACCGGCGCTGCTGGAGCGGCTGCGGAGCTGGGCGTGACCGCGGTCACTGTAC 584
 QY 370 CACTGGATCTGCTCAGCGCTTCAAGATCGCTATGAGAGCTGCTCAACCTGGAAGAG 429
 Db 585 CACTGGACCTGCGCCAGCGCTGACAGAGCGCTACGCGCGCTGGGCAACCGCGCTTG 644
 QY 430 GTCCAGCTGAGCTTTGAGCGGCTATGAGAGGTGCTTTGACGTTTGGGAGCCGAGTC 489
 Db 645 GCGGACC---ACTTCAGGAGTTACGGGAGCTTGTCTTCCGCACTTCCGCGGTCAAGTC 701
 QY 490 CAGAACTGATCAACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGGC 549
 Db 702 AAGTACTGATCAACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGGC 761
 QY 550 AGCAAGCCCGGCGAGAGCAGATTTACAGCACTTCAACCGAGGCAACACTGCCACT 609
 Db 762 CGCTGGCCCCCGGCAATCCGGGGGAGCCCGGCTC----- 797
 QY 610 GAGCGGTGCTCGCTGGAAAGGCCAGATCGGCTCAACGAGGAGTAC 669
 Db 798 ---GGTACCTGGGGGAGCAACCTCTCTGCTGCTGCAAGGCTTGGCATCTCTAC 854
 QY 670 AGCAGGAGCTTTGGCCCCGCGCAAAAGGCCAGATCGGCTCAACGAGGAGTAC 729
 Db 855 AATCTCTTCTTCCTCCACTCAGGAGGTGAGGTGCTCATGCTTCAAGCTTCACCTGG 914
 QY 730 TATGAGCCTTGGGACAGCAATGAGCTTCGGGCAAGAGGCTGTCGACGAGTGA 789
 Db 915 ATCAATCTCGAAGAAATGACCGACCAAGCATCAAAAGATGCAAAATCTC---TGGAC 971
 QY 790 TTTCACTTGGCTGTGTTGGCAATCCCATCTTCTTGAAGAAGCTTCCAGAGAGCATG 849
 Db 972 TTTGTACTAGGTGTGTTGGCAACCCGTATTTATGATGCTATCCCGAGAGCATG 1031
 QY 850 AAGAAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGCCATCTCAAT 909
 Db 1032 AAGATTAACCTTTCACTATTTCTGCTGATTTTACTGAATCTGAAGAAAGTTATCAAAA 1091
 QY 910 G 910
 Db 1092 G 1092

Dh 482 ATATAGTACCACAAATTGCTTTCTATGCTGCTGTGTAACCTATACAGAGA 541
Qy 677 ACTTTCGCCCTCGCAAAAGGGCCAGATCGGCATCTGCTCAAGGGAGCTACTATGAGC 736
Db 542 AATATCAGGCAATATCAAAAGGGCCAGATAGGATTAACATAGTACTTATGATGATTC 601
Qy 737 CTTGGGACAGCAATGACCTTCGAGCAAGAGCTGTGAGCGACGATGGAATTTGACA 796
Db 602 CCT---ACTCCAAATTGAAAGCGACAGAGATGACAGCAAGCCCTGATTTCAATGT 658
Qy 797 TTGGCTGCTTTGCCAATCCCATCTTTCTTGAAGAAGACTATCCAGAGCATGAAGAGC 856
Db 659 ATGATGTGTTATTAAGCCATTAAGCTT---TGCTGATATTCACAAAAGCATGCGTAGAC 715
Qy 857 AGCTGGCGAGAGGCTTCACGCCCTCACTCCCGCGGA 893
Db 716 TCGTTGTTAAAGGTTTACCAAGGTTCACTAAAGACA 752

RESULT 9

US-09-122-230-8
; Sequence 8, Application US/09122230A
; Patent No. 5973228
; GENERAL INFORMATION:
; APPLICANT: Carlson, et al.
; TITLE OF INVENTION: Coniferin Beta Glucosidase cDNA for Modifying Lignin
; TITLE OF INVENTION: Composition in Plants
; FILE REFERENCE: 50532
; CURRENT APPLICATION NUMBER: US/09/122,230A
; CURRENT FILING DATE: 1998-07-23
; EARLIER APPLICATION NUMBER: U.S. 60/053,566
; EARLIER FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Pinus contorta
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1542)
US-09-122-230-8

Query Match 8.6%; Score 124.8; DB 2; Length 1542;
Best Local Similarity 54.4%; Pred. No. 1e-22;
Matches 299; Conservative 0; Mismatches 242; Indels 9; Gaps 2;

Qy 20 TGCCCAACGACTTGAATGGGGCTTCGCAACGGCCGCTACAGATGGAAGGCGCGTCA 79
Db 89 TCCCTCAGATTTCATGTTGGGACACAGCTCTTCAGGGTATCAATAGAGAGACAGTCC 148
Qy 80 AAGAAGGTGCGCGCGCGCGCTTCATCTGGGACACATGCTGCACTGACCTGAGACATCGGCA 139
Db 149 GAGAGAGTGGCAAGGGTCTTAGCAGATGGGACGCTTAAACATATATCTGTGAAATAA 208
Qy 140 CCAAGCGGCAACGGGAGTGGCTTGCATCACTACACCGCTAGATGAGAGACTTTG 199
Db 209 AAGTATACAGAAAGAGAGAGCTGCGAGTCCAGCAATATCAAGATATATGGAAGATATCG 268
Qy 200 ATCTCTGACCAAGTACGGGCGCAAGGCTTACCGCTTCTCTGCTGTGCGGATCA 259
Db 269 AGCTTATGGCTTCACTTGGACTAGATGCTATATGATTTCTCATATCTCGTCTGAAATCC 328
Qy 260 TTCCCTCGCGGCGAGCGCTGATCCGTCACAGAGAGGAATGAGTTTACAGCAAAAC 319
Db 329 TTCCAGAGGAAG-----AGGTAAATTAACATAGGCTGGGATGAATATTAACAATATC 382
Qy 320 TGAATGACGCGCTGTGAGGCGGGTATCACGCGCTTGGGTGACTTTGACCACTGGGATC 379
Db 383 TGAATGACGCTTCTTCTGCAAAATGGGATCCAGCGCTTGTGACATTTGTTCCATTTGCAATC 442
Qy 380 TGCTCAGGCGCTTACGATGCTATGAGGCTGCTCAACGTGAAAGAGTCCAGCTGCG 439

Dh 443 TTCCCAAGCACTTGAAGACTCTATGCGGAGATGCGTAGCTCTCAATAATTA---CG 499
Qy 440 ACTTTCAGGCTATGCGAGGTGAGCTTGAAGCTTTTGGGAGCCGAGTCCAGAACTGGA 499
Db 500 ACTTCAGGCTTATGAGAGATTTGCTTCGGGCACTTCGATGACCGTCAATAATTTGGG 559
Qy 500 TCAACATCAACGAGCCCTGATTCAGGCCATCTATGAGATATGCAACCGAGCAACGCC 559
Db 560 CGACAGTGAACGAGCCAAATCTGTTGTGCGTTGGATACCGCGTGGAAATTTTCAC 619
Qy 560 CGGCGAGGAG 569
Db 620 CGACGAGGTG 629

RESULT 10

US-09-122-230-6
; Sequence 6, Application US/09122230A
; Patent No. 5973228
; GENERAL INFORMATION:
; APPLICANT: Carlson, et al.
; TITLE OF INVENTION: Coniferin Beta Glucosidase cDNA for Modifying Lignin
; TITLE OF INVENTION: Composition in Plants
; FILE REFERENCE: 50532
; CURRENT APPLICATION NUMBER: US/09/122,230A
; CURRENT FILING DATE: 1998-07-23
; EARLIER APPLICATION NUMBER: U.S. 60/053,566
; EARLIER FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1909
; TYPE: DNA
; ORGANISM: Pinus contorta
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (183)..(1724)
US-09-122-230-6

Query Match 8.6%; Score 124.8; DB 2; Length 1909;
Best Local Similarity 54.4%; Pred. No. 1e-22;
Matches 299; Conservative 0; Mismatches 242; Indels 9; Gaps 2;

Qy 20 TGCCCAACGACTTGAATGGGGCTTCGCAACGGCCGCTACAGATGGAAGGCGCGTCA 79
Db 271 TCCCTCAGATTTCATGTTGGGACACAGCTCTTCAGGGTATCAATAGAGAGACAGTCC 330
Qy 80 AAGAAGGTGCGCGCGCGCGCTTCATCTGGGACACATGCTGCACTGACCTGAGACATCGGCA 139
Db 331 GAGAGAGTGGCAAGGGTCTTAGCAGATGGGACGCTTAAACATATATCTGTGAAATAA 390
Qy 140 CCAAGCGGCAACGGGAGTGGCTTGCATCACTACACCGCTAGATGAGAGACTTTG 199
Db 391 AAGTATACAGAAAGAGAGAGCTGCGAGTCCAGCAATATCAAGATATATGGAAGATATCG 450
Qy 200 ATCTCTGACCAAGTACGGGCGCAAGGCTTACCGCTTCTCTGCTGTGCGGATCA 259
Db 451 AGCTTATGGCTTCACTTGGACTAGATGCTATATGATTTCTCATATCTCGTCTGAAATCC 510
Qy 260 TTCCCTCGCGGCGAGCTGATCCGTCACAGAGAGGAATGAGTTTACAGCAAAAC 319
Db 511 TTCCAGAGGAAG-----AGGTAAATTAACATAGGCTGGGATGAATATTAACAATATC 564
Qy 320 TGAATGACGCGCTGTGAGGCGGGTATCACGCGCTTGGGTGACTTTGACCACTGGGATC 379
Db 565 TGAATGACGCTTCTTCTGCAAAATGGGATCCAGCGCTTGTGACATTTGTTCCATTTGCAATC 624
Qy 380 TGCTCAGGCGCTTACGATGCTATGAGGCTGCTCAACGTGGAAGAGTCCAGCTGCG 439
Db 625 TTCCCAAGCACTTGAAGACTCTATGCGGAGATGCGTCAATTAATTA---CG 681
Qy 440 ACTTTCAGGCTATGCGAGGTGCTTGAACGTTTGGGAGCCGAGTCCAGAACTGGA 499

Db 662 ACTTGGAAGCCTATGACAGAGATTGCTTCCGGGCAATCGGTGACCGGTGCAATATATGCG 741
QY 500 TCACCATCAACGAMCCCTGATTCAGGCACTATGATGATGACCGGCGCAACGCC 559
Db 742 CGACAGTAGAAGACGCAATCTGTTGTGCGCTGGATACCGCTCGAATATTTCCAC 801
QY 560 CGGCGAGAG 569
Db 802 CGACGAGGTG 811

RESULT 11
US-08-688-988-9
; Sequence 9, Application US/08688988B
; Patent No. 6096545
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Daniel D.
; APPLICANT: Malboobi, Mohammad A.
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
; FILE REFERENCE: PPL96-03
; CURRENT APPLICATION NUMBER: US/08/688,988B
; CURRENT FILING DATE: 1996-07-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1829
; TYPE: DNA
; ORGANISM: Arabidopsis Thaliana
US-08-688-988-9

Query Match 8.2%; Score 118.6; DB 3; Length 1829;
Best Local Similarity 53.6%; Pred. No. 4.3e-21;
Matches 294; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

QY 18 TCTGCCCAACGACTTGAATGGGCTTGCACACGCGCCCTACAGATCGAAG---CGC 74
Db 159 TTTCCCTGAAGGTTTATTTTGGACGCGTACTGCGCATACAGAGGATCCAGGTTGCA 218
QY 75 CGTCAAGAGGTGCGCGCGCGCCGCTCCATCTGGGACACGTACTGCCACTTGAGCATC 134
Db 219 TTTATGAAACTTGTGTGGACCGCTTATGGGACATCTACTGTGAATATTCAGA 278
QY 135 GCGCACCAACGCGCCCAACGCGCATGTGGCTTGGATCCTACACCGCTACGATGAGA 194
Db 279 GAGGTGCAATACATACACGCGCATGTGGCGGTGATTTCTTCATGCTATAGAGA 338
QY 195 CTTGATCTCTTGAACCAAGTACGGGCGCAAGCGCTTCTCTTGTGCTGCG 254
Db 339 TATCCACTATGAAAGATCTAAACACAGACGCTTTTGAATGTCTATCGATGGCCAAG 398
QY 255 GATCATTCCTCGCGCGCGAGGTGATCCCGTCAACGAGAGGAAATGATTTTACAG 314
Db 399 AATATTTCTCATGGAGAAAGAGAAAGAGTGAAGTCAAGCTGTGTGCAATTTTCCA 458
QY 315 CAATCTATTTAAGCCCTGTGTGAGCGGGGTATACCGCTTGGGTACTTTGACCACTG 374
Db 459 CGACCTATGACGACGCTCATATAAAATGTATTAATCTCAATCTGTTTTCCTG 518
QY 375 GGATCTGCTCAGGCGCTTCACATCGCTATGAGGCGGTGCTCAAGCTGGAAGGTCGA 434
Db 519 GGAACCTCCACAAATTTAGAAATGAAATATGAGGCGCTTTTAAAGC---GAAAGATTGT 575
QY 435 GCTGGACTTTGAGCGGTATGCGAGGTTGTGCTTGAACGTTTGGGAGCGAGTCCAGAA 494
Db 576 GAGAGATTTCCGAGAGTATGAGATTTGTTTCCAGAAATAGCGTGAAGAAAGTGAACA 635
QY 495 CTGGATACCATCAACGAMCCCTGATTCAGGCACTATGATGATATCCACCGGACGAA 554
Db 636 TTGGATCACTTTCAATAGGACATGAGGTTTCTCGACGCTGTGATGACGTAAGCAAAA 695
QY 555 CGCCCGGAG 563
Db 696 GGACCTGG 704

RESULT 12
US-08-400-275-17
; Sequence 17, Application US/08400275
; Patent No. 5668295
; GENERAL INFORMATION:
; APPLICANT: Mahab, Samir Z.
; APPLICANT: Mahab, Vedpal S.
; TITLE OF INVENTION: PUTRESCINE N-METHYLTRANSFERASE
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING PUTRESCINE
; TITLE OF INVENTION: N-METHYLTRANSFERASE, AND TRANSGENIC TOBACCO PLANTS WITH
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Ave. of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10020

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,275
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,681
; FILING DATE:
; APPLICATION NUMBER: US 07/613,160
; FILING DATE: 14-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
; REFERENCE/DOCKET NUMBER: PM-1696
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090

; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1545 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-400-275-17

Query Match 7.7%; Score 111.4; DB 1; Length 1545;
Best Local Similarity 49.9%; Pred. No. 2.9e-19;
Matches 364; Conservative 0; Mismatches 356; Indels 9; Gaps 3;

QY 14 TAGCTGCGCAACGACTTGAATGGGCGTTGCAACGCGCCCTACAGATCGAAGCG 73
Db 2 TGCGCTTTCTTCTATATTTCTTGTGCGGACGCTTTCATATTAACAGATGAAGAG 61
QY 74 CCGTCAAGAGATGCGCGCGCGCCGCTCCATCTGGGACACGTATGCACTGAGCAAT 133
Db 62 CTTTCTCGATATGGAAGAGGCTCAGCAACTGGAGCGTTTAAACCATGAAGCTGGTC 121
QY 134 CGGCAACCAACGCGCAACGCGCATGTGGCTTGGATCTACTACACCGCTACGATGAG 193
Db 122 ATGTTAAGATGAAGACATGAGATGTGGCTGTGATCATACCATGTTATTTGAGG 181
QY 194 ACTTGATCTCTTGAACCAAGTACGGCGCAAGGCGCTACCGCTTCTCTTGTGCTGCG 253
Db 182 ACATCAACTCATGGCAATATGGGTGTGAATAGCTTGTCTTATTCATGAGGC-- 239

Oy	254	GGATATTTTCCCTCGGCGGCGAGGCTGATCCCGTCAAGAGAGAGGAAATTGAGTTTAA	313
Db	240	-AAGAAITCTGCCCAAGGAAATATTGGAGAAATTAAATTGGCCGAAATTTGACACTACA	298
Oy	314	GCAAACTGATTGACGCCCTGTTGAGCGGGGATATCAGCGCTTGCGGTACTTTGTACACT	373
Db	299	GTAACCTGATTGATCAGCTCTTACAGAAAGGATCAGCCGTTTGTCACTTAAACATTT	358
Oy	374	GGGATCTGCTTCAGCGCGCTTCAAGATCGTATGAGGCGTGGCTCAACGTGGAAGAGTCC	433
Db	359	TTGAAATACCAACAAACTTGGAGCAAGATATGTGTGGCTAA--GTTCACAATAC	415
Oy	434	AGCTGACTTTGAGGCGTATGCGAGGTTGTGCTTGAACGTTTTGGGGAACGAGTCAGA	493
Db	416	GGATGATTTTCAGCTATTTCGAAACATATGCTTCAAAATACTTGGGAGATAGATTAAAT	475
Oy	494	ACTGATATCAACATCAACGAMCCCTGATTCAGGCGCATCTATGATATGCCACCGCAGCA	553
Db	476	ACTGGGTAAACATGATGATAGAGCTTAACTTCGTGCCCTTAATGTGCTATTAAGATGGACTTT	535
Oy	554	ACGCCCCCGGAGAGAGC--AGCATTTAACAAACACTCCACGAGGGCAACACTGCCACTG	610
Db	536	GCCCTCCAACTGATGCTCTGTGTTAATTTGGGAATTGTAGTGTGGGATTCAGAAAGGG	595
Oy	611	AGCGTGGCTCGCTGGAAGGCCCAAGATATGAGCATATGCCGCGCGCGTGGCGCTTACA	670
Db	596	AACCTTTGATGCACTACCAATATGATCTTATCTATGACAGATGTGTGCATGACATTTAAC	655
Oy	671	GCAGGGACTTTGCGCCCTCGCAAAAGGCGCAATCGGATCTTGCATCAACGCGCACTACT	730
Db	656	GCACCAAGATATCAGAAAGTCAAGAGGCAATGATTGCACTTATGCGGTTTCGATTGCT	715
Oy	731	ATGAGCCCT	739
Db	716	ATGAACCGT	724

```

RESULT 13
US-09-134-000C-2890
; Sequence 2890, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2890
; LENGTH: 1458
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; US-09-134-000C-2890

```

Query Match	7.2%	Score 104.8	DB 4	Length 1458
Best Local Similarity	53.4%	Pred. No. 1.4e-17		
Matches 267	Conservative 1	Mismatches 223	Indels 9	Gaps 2

QY	15	AGCTCTGCCCAACGACTTTGATGGGGCTTGGCAACGGCCGCTACGATCGAAGCGC	74
Db	45	AGCATTTCCGAAGATTTTATATGGGGCTCTGCTTCGGCAGCTTATCAAGTCGAAGTGC	104
QY	75	CGTCAAGAAGGTGGCCGGCCGCTCCATCTGGGACAGTACTGCCACTGGAGCCATC	134
Db	105	TTGGAATGAAAGCGCAAGGCGCATCTGTTGGATGAATTTGTTGGCTACAGAGAA	164
QY	135	GCGCAACCAAGCGCGCCCAACGGCGATGTGTGGCTTGGCATCACTAACCGCTAGATGAGA	194
Db	165	AACGTTCCAAAGACCACTGGCGATTTTAGCCGTCGATCATTTATCATCTGTTCCAAAGAGA	224

Oy	195	CTTTGATCTCTGAACAGTAAGCGCAAAAGCCTAACGGCTTCTCCTTGTCGTGGCGGG	254
Db	225	TGTGCCTTTAATGAAAACAAGAAGTTTAAAAAGCATATGTTTTCTGATTGCATGGACAG	284
Oy	255	GATCATTTCCCCTCGGCGGCGAGGCTGATATCCGCTCAACGAGAGGAAATTGAGTTTTCAG	314
Db	285	AATTTTACC-----TGAAGGCCGTGTGCGAGTAAATCAGCGGGACTGAATTTTATTTC	338
Oy	315	CAAACGTATTGACGGCCCTGTGAGGCGGGGATACAGCCTTGGGTGACTTTGTACCACTG	374
Db	339	AGATTTAATTATGATGATTATTAAGCGCAGGAATTGAACCAATGTTAATTCTTGATCATTTG	398
Oy	375	GGATTCGCTCAGGCGCCTTCAAGATCGCTTAGAGGCTGGCTCAACGTGGAAGAGGTCCA	434
Db	399	GGATTTCGCACCTGTTTGGCAAAAAAGAAATATGTGGCTGG---GAATCAAGAAAAATTAT	455
Oy	435	GCTGGACTTTAGACCGGTAATCGGAGTTGTGCTTTGAACGTTTGGGGAACCGAGTCCAGAA	494
Db	456	TGATGATTTTBTGGCCTATGGGAAAATTTTATTGACGCTTTCGTTGGCAAGATTCGCTTA	515
Oy	495	CTGATCACCATCAACGAMC 514	
Db	516	TTGGATTATGTTAAATGAAC 535	

```

RESULT 14
US-09-107-532A-1268
Sequence 1268, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Walham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ. ID NO. 1268:
SEQUENCE CHARACTERISTICS:
LENGTH: 1437 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

```

```

FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1437
SEQUENCE DESCRIPTION: SEQ ID NO: 1268:
US-09-107-532A-1268

```

```

Query Match      7.0%; Score 101.4; DB 4; Length 1437;
Best Local Similarity 53.1%; Pred. No. 1.1e-16;
Matches 263; Conservative 1; Mismatches 222; Indels 9; Gaps 2;

```

```

QY 20 TGCCTACGACTTGAATGAGGCTTTCGACGCGCGCTTACCAATCGAAGCGCGCTCA 79
DB 26 TTCCTAAAGATTCTGTGGGCTCTGTCTTCGACGCTATCAAGTAAAGAGCTTGGC 85
QY 80 AAGAAGTGGCGCGCGCGCTTCATCTGGGACAGTACTGCGCACTGAGAGCATCGCCCA 139
DB 86 AAGAAGATGGGAAAGAGAAATCAAGCTGGGATCGATTGTACGATACCTGAGAAAGCAT 145
QY 140 CCAACGCGCGCCACGCGATGCTGCTGATCACTACCACTACCGCTACGATGAGACTTTG 199
DB 146 TTTAAGGACGACGAGTATTTGGGAGTAGACACTATCATCTTTAATAAGAAATATG 205
QY 200 ATCTCTTGACCAAGTACGCGCCAAAGGCTTACCGCTTCTCTGTCTGTGCGGATCA 259
DB 206 CTTTATGAAAGAACAAAGGTCTGAAAGCTTACCGCTTCTGTAGCTTGAACGAGATT 265
QY 260 TTCCCTCGCGCGCGCTGATCCCTCAACGAGAGGAATTGATTTTACAGCAAC 319
DB 266 TTCCCTAATG-----GACGAGAGAGGATCAACAGCTGGGCTGCTTTTATGAGAT 319
QY 320 TGATTCAGCGCTTGTGAGCGGCGGTATCACGCTTGGGTGACTTTTACAGCTGGATC 379
DB 320 TGATTCAGGATTTGATTTGAAACGATTTGAACCTATTTGACTTTATATATGAGATC 379
QY 380 TGCCTCAGCGCTTACCATGCTATGAGGCTGCTCAACGTGAAAGGTCCAGCTGG 439
DB 380 TTCCCAAGCATTTGCAAGATGATATATATGTTGG---GAATCACGTGATCATTTGAG 436
QY 440 ACTTTGAGCGGATGCGAGGTGCTTGTGAACGTTTGGGGACCGAGTCCAGACTGGA 499
DB 437 ATTTCATCTATATGCAAGAAACATTATTTGAGGATTCGCGGGAAGTCCATTACTGA 496
QY 500 TCACCATCAGCAGMC 514
DB 497 TTAGTTGATGATGAC 511

```

```

RESULT 15
US-09-489-039A-1053
; Sequence 1053, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KIEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1053
; LENGTH: 1410
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1053

```

```

Query Match      6.6%; Score 95.6; DB 4; Length 1410;
Best Local Similarity 51.9%; Pred. No. 3.3e-15;
Matches 302; Conservative 1; Mismatches 255; Indels 24; Gaps 3;

```

```

QY 8 AGTCCTAGCTCTGCCAAGACTTTGAATGAGGCTTGGCAACGCGCGCTTACCAAGATCG 67

```

```

DB 14 AGATGAAACATTTCCGACCGATTTTTTATGAGGCGCGCGGACCGCGCAATCAGGTGG 73
QY 68 AAGGCGCGCTTCAAGAAAGATGAGCGCGCGCTTCATCTGGGACAGCTATGCGACC--- 124
DB 74 AAGGTGCTTATCTGGAAGATGCAAAAGCTTATTCACCTCCGATGTCAGACCCGACGGC 133
QY 125 -----TGAACCATCGCGCACCAACGCGCGCAACGCGGATGAGCTTGGC 169
DB 134 TTTTCGCGAAGTGTGACGCGGTACCCGCGACAGCGCGCATTAAGATGTTGCCATTCG 193
QY 170 ATCACTACGACCGCTACGATGAGACTTTTATCTTTGACCAATACGAGGCAAGGCTT 229
DB 194 ATTTTACATCTGCTATCCGAAATATGCGCTGTTTGCAGATGAGCTTTTAACTGCC 253
QY 230 ACGGCTCTCTGCTGTGCTGCGGATCATTCCTCGCGGACGAGCTGATCCGCTCA 289
DB 254 TGCAGTCTCTATGCTTGAACCGCATCTAACCCAAACGCGGACGACGCGGAGCC---GA 310
QY 290 ACGAGAGGGAATTTGATTTTACAGCAACTGATTCAGCGCTGTGAGCGCGGATATCA 349
DB 311 ACGAAGCGGCGCTGCGTTTACGACAAAGCTTTCAGAGATGCGAGAACACAAACATCA 370
QY 350 CGCTTGGGTGACTTTTATACCATGAGATCTGCTCAGGCGCTTACAGATCGCTATGAG 409
DB 371 CCCGCTGTGACGCTGTGCAATTATGAAATGCCATGCGCTGTGTAATAAACTATAGCG 430
QY 410 GCTGCTCAACGTTGAGAGGATCCAGCTTGAAGCTTGTGAGCGGTATGCGAGGTGTGCTTTG 469
DB 431 GCTGGGCGACCGGAGAGTATGCTT---CTTTGAGCGCTTACGCGGAGCGTGTCA 487
QY 470 AACGTTTGGGACCGAGTCCAGAACTGATTCACATCAACGAMCCTGATTCAGGCCA 529
DB 488 GCGCTATTAATAAATTAAGTCAAACTGTGCTGACTTTTAAAGAAATCAATATGTGCTGC 547
QY 530 TCTATGATATGCCACCGGACGACAGCCCGCGGACGAGACA 571
DB 548 ATGCCCGATGACCGGCGTGGCTGCGCGCGCGGACGACGA 589

```

```

Search completed: March 28, 2004, 20:56:57
Job time : 109.557 secs

```


QY	181	CCCTACGATGAGACCTTTATCTCTTGACCAAGTACGCGCAAAAGCCTTACCGCTTCTCC	240
Dp	181	CGCTACGATGAGACCTTTATCTCTTGACCAAGTACGCGCAAAAGCCTTACCGCTTCTCC	240
QY	241	TGTGTGTGGTCCGCGATCATTCGCCCTCGCGGAGGCTGGATTCGCCGTCAACGAGAGGGGA	300
Dp	241	TGTGTGTGGTCCGCGATCATTCGCCCTCGCGGAGGCTGGATTCGCCGTCAACGAGAGGGGA	300
QY	301	ATTGAGCTTTACAGCAAACTGATGACGCTCTGTTGAGGCGGGGATGTCACGCGCTTGGGTG	360
Dp	301	ATTGAGCTTTACAGCAAACTGATGACGCTCTGTTGAGGCGGGGATGTCACGCGCTTGGGTG	360
QY	361	ACTTGTATACCACTGGGATCTGCCTCAGAGGCTTCAAGATGGCTATGGAAGGCTGCTCAAC	420
Dp	361	ACTTGTATACCACTGGGATCTGCCTCAGAGGCTTCAAGATGGCTATGGAAGGCTGCTCAAC	420
QY	421	GTGGAAGAGGTCCAGCTGACTTTGACCGGTAATGGCAGTTGTCTTTGAACGTTTGGG	480
Dp	421	GTGGAAGAGGTCCAGCTGACTTTGACCGGTAATGGCAGTTGTCTTTGAACGTTTGGG	480
QY	481	GACCCAGTCCAGAATCGATATCAACATTCACGAGCCCTGGATTCAGGCCATCTATGATAT	540
Dp	481	GACCCAGTCCAGAATCGATATCAACATTCACGAGCCCTGGATTCAGGCCATCTATGATAT	540
QY	541	GCCACCGGACGACACGCCCCGGGCGAGAGCAGACTTAAACAGCATCTCCACCGAGGGCAAC	600
Dp	541	GCCACCGGACGACACGCCCCGGGCGAGAGCAGACTTAAACAGCATCTCCACCGAGGGCAAC	600
QY	601	ACTGCGACTTGAAGCCGTGGCTCGCTGGAAAAGGCCAGATCATGAGCCATGCCCGCCGTG	660
Dp	601	ACTGCGACTTGAAGCCGTGGCTCGCTGGAAAAGGCCAGATCATGAGCCATGCCCGCCGTG	660
QY	661	GCCGTCTACAGCAGGAGACTTTCGCCCTCGCAAAAGGCGCAGATCGGCATCTCGCTCAAC	720
Dp	661	GCCGTCTACAGCAGGAGACTTTCGCCCTCGCAAAAGGCGCAGATCGGCATCTCGCTCAAC	720
QY	721	GCGCGTCTATGAGCCCTCGGAGCAACATGAGCTCGGGACAAAGAGGCTGCTGAGCGA	780
Dp	721	GCGCGTCTATGAGCCCTCGGAGCAACATGAGCTCGGGACAAAGAGGCTGCTGAGCGA	780
QY	781	CGGATGGAATTTCAATTGGCTGGTTTGCAATCCCATCTTTGAGAAGAGCAATACCA	840
Dp	781	CGGATGGAATTTCAATTGGCTGGTTTGCAATCCCATCTTTGAGAAGAGCAATACCA	840
QY	841	GAGAGCATGAAAGAACACACTGGCGAGAGGCTTCCAGCCTTCACCTCCGCGGACTTTGCC	900
Dp	841	GAGAGCATGAAAGAACACACTGGCGAGAGGCTTCCAGCCTTCACCTCCGCGGACTTTGCC	900
QY	901	ATCTCTCATATGCGGAGAGACCGGACTTTCACCGGATGAAATTTCAACATCCCATGTTCCG	960
Dp	901	ATCTCTCATATGCGGAGAGACCGGACTTTCACCGGATGAAATTTCAACATCCCATGTTCCG	960
QY	961	CGCCACCTTAAACGCTGCCGCTCCCGAGACGGAATATCTCGGCGCCATCATAGAGCACAG	1020
Dp	961	CGCCACCTTAAACGCTGCCGCTCCCGAGACGGAATATCTCGGCGCCATCATAGAGCACAG	1020
QY	1021	GAGAAATAAGACGGGACGCCGCTGGGCGAGAGAGCGGCGCTCGCTGGCTGCGCTCTGC	1080
Dp	1021	GAGAAATAAGACGGGACGCCGCTGGGCGAGAGAGCGGCGCTCGCTGGCTGCGCTCTGC	1080
QY	1081	CCGGACATGTTCCGGAAGCATCTGCGCCGGGTGTACCGGCTGTACGGCAACCATCTAC	1140
Dp	1081	CCGGACATGTTCCGGAAGCATCTGCGCCGGGTGTACCGGCTGTACGGCAACCATCTAC	1140
QY	1141	ATCACCGAGAAACGATGCGCTGCTGGAGAGGAAATGACGTGCGAGAGAGGCGCTC	1200
Dp	1141	ATCACCGAGAAACGATGCGCTGCTGGAGAGGAAATGACGTGCGAGAGAGGCGCTC	1200
QY	1201	AACACACCCCTTCCGCATCCGCTTACCTTGAATCTGACCTGCACTTTCAGAGGCAAT	1266
Dp	1201	AACACACCCCTTCCGCATCCGCTTACCTTGAATCTGACCTGCACTTTCAGAGGCAAT	1266
QY	1261	ACCCAGAGCGGCGTGTCTCAAGGGGTACTTTGGGTGGGCTGCTGATACCTTGGAA	1320

DB	1261	ACCACGAGCGCGTGTGTGTGTCAGTAAAGGGGTA	CTTTGTGGGCGTTTGCTGATTA	CTTGGAA	1320
QY	1331	TGTCAGATGGGTACGGAACCAAGATTTGGCGCTCA	ACGTTCA	CAGACTA	CAACACCTCTAAG 1380
DB	1321	TGTCAGATGGGTACGGAACCAAGATTTGGCGCTCA	ACGTTCA	CAGACTA	CAACACCTCTAAG 1380
QY	1381	CGACGCCCAAGAAAGTCTGCCCTGTCTTCAAGACA	TGTTTGGGCCCGGACAGAGGTT		1440
DB	1381	CGACGCCCAAGAAAGTCTGCCCTGTCTTCAAGACA	TGTTTGGGCCCGGACAGAGGTT		1440
QY	1441	AAAGTGGGCGCATAA	1455		
DB	1441	AAAGTGGGCGCATAA	1455		
RESULT 2					
US-10-026-140-1					
Sequence 1, Application US/10026140					
Publication No. US20030114330A1					
GENERAL INFORMATION:					
APPLICANT: Dunn-Coleman, Nigel					
APPLICANT: Goedegebuur, Frits					
APPLICANT: Ward, Michael					
APPLICANT: Yao, Jian					
TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids					
TITLE OF INVENTION: Encoding the Same					
FILE REFERENCE: GC697					
CURRENT APPLICATION NUMBER: US/10/026.140					
CURRENT FILING DATE: 2002-03-26					
NUMBER OF SEQ ID NOS: 3					
SOFTWARE: FastSeq for Windows Version 4.0					
SEQ ID NO 1					
LENGTH: 1991					
TYPE: DNA					
ORGANISM: Trichoderma reesei					
US-10-026-140-1					
Query Match					
Best Local Similarity 99.9%; Score 1454.2; DB 14; Length 1991;					
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	ATGCCGAGTGTGCTAGCTTGTGCCCAACGACTTTGAAT	GGGGCTTGC	CAACGGCCGCTAC	60
DB	62	ATGCCGAGTGTGCTAGCTTGTGCCCAACGACTTTGAAT	GGGGCTTGC	CAACGGCCGCTAC	121
QY	61	CAGATCGAAGCGCGCTCAAAAGAGTGGCGCGCGCCGCT	CACTTTGGGACAGTACTGC		120
DB	122	CAGATCGAAGCGCGCTCAAAAGAGTGGCGCGCGCCGCT	CACTTTGGGACAGTACTGC		181
QY	121	CACCTGAGGCATTCGCGCAACCAACGGCGCCACGGCGAT	TGGCTTGCATCACTACAC		180
DB	182	CACCTGAGGCATTCGCGCAACCAACGGCGCCACGGCGAT	TGGCTTGCATCACTACAC		241
QY	181	CGCTACGATAGGACTTTGATTTCTTTGACCAAGTAGCGG	CCCAAAAGGCTTACCGTTCTCC		240
DB	242	CGCTACGATAGGACTTTGATTTCTTTGACCAAGTAGCGG	CCCAAAAGGCTTACCGTTCTCC		301
QY	241	TTGTCTGTCTCGCGGATCATTTCCCTTGGCGGACAGCT	GTGATCCCGTCAACGAGAGGA		300
DB	302	TTGTCTGTCTCGCGGATCATTTCCCTTGGCGGACAGCT	GTGATCCCGTCAACGAGAGGA		361
QY	301	ATTGAGTTTACAGCAAACTGATTGACGCCCTGTGAGG	CGGGGATACGCTTTGGGTG		360
DB	362	ATTGAGTTTACAGCAAACTGATTGACGCCCTGTGAGG	CGGGGATACGCTTTGGGTG		421
QY	361	ACTTTGACACATGGGATTCGCGTCAAGGCGCTTCA	CGATGCGTATGAGAGGCTTGCAAC		420
DB	422	ACTTTGACACATGGGATTCGCGTCAAGGCGCTTCA	CGATGCGTATGAGAGGCTTGCAAC		481
QY	421	GTGGAAGAGGTCCAGCTGACCTTTGAAGCGGTATG	CGAGTGTGTCTTTGAACGTTTGGG		480
DB	482	GTGGAAGAGGTCCAGCTGACCTTTGAAGCGGTATG	CGAGTGTGTCTTTGAACGTTTGGG		541

481 GACGAGTCCAGAACTGGATTCACATCAAGAACCTTGATTCAGGCATCTATGATAT 540
482 GACCGAGTCCAGAACTGGATTCACATCAAGAACCTTGATTCAGGCATCTATGATAT 601
541 GCCACCGGACAGCAAGCCCGGAGGAGAGCACTTAAACAAGCACTCCACGAGGGCAAC 600
602 GCCACCGGACAGCAAGCCCGGAGGAGAGCACTTAAACAAGCACTCCACGAGGGCAAC 661
601 ACTGCACTAGAGCGGTGCTGCTGCGGAAGCCCAAGTCAATGAGCCATGCGCGCGCTG 660
662 ACTGCACTAGAGCGGTGCTGCTGCGGAAGCCCAAGTCAATGAGCCATGCGCGCGCTG 721
661 GCGCTCAACAGAGGAGCTTCCGCTCCGCAAAAAGGAGAGATCGGATCTCGCTCAAC 720
722 GCGCTCAACAGAGGAGCTTCCGCTCCGCAAAAAGGAGAGATCGGATCTCGCTCAAC 781
721 GCGGACTACTATGAGCCCTGAGAGCAATGAGCTCGGAGCAAGAGGCTGCTGAGCGA 780
782 GCGGACTACTATGAGCCCTGAGAGCAATGAGCTCGGAGCAAGAGGCTGCTGAGCGA 841
781 CGGATGGAATTTTCAATTTGGTGGTTTGGCAATCCCATCTTTTGAAGAAAGACTATCCA 840
842 CGGATGGAATTTTCAATTTGGTGGTTTGGCAATCCCATCTTTTGAAGAAAGACTATCCA 901
841 GAGAGCATGAAGAAGAGCTGGGAGAGGCTTCAGCCCTCACTCCGCGAGACTTTGGC 900
902 GAGAGCATGAAGAAGAGCTGGGAGAGGCTTCAGCCCTCACTCCGCGAGACTTTGGC 961
901 ATCTCTAATGCGGAGAGAGCACTTCTACGCGATGAATTAATCAACATCCAGTTGGCG 960
962 ATCTCTAATGCGGAGAGAGCACTTCTACGCGATGAATTAATCAACATCCAGTTGGCG 1021
961 CGGCACTTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
1022 CGGCACTTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1081
1021 GAGAAATGAAG 1080
1082 GAGAAATGAAG 1141
1081 CGGAACTGTTCCGAG 1140
1142 CGGAACTGTTCCGAG 1201
1141 ATCAACGAG 1200
1202 ATCAACGAG 1261
1201 AACGAG 1260
1262 AACGAG 1321
1261 ACCGAG 1320
1322 ACCGAG 1381
1321 TGGTCAATGAG 1380
1382 TGGTCAATGAG 1441
1381 CGGAG 1440
1442 CGGAG 1501
1441 AAGTGGCGGAGATTA 1455
1502 AAGTGGCGGAGATTA 1516

RESULT 3
US-10-369-493-37092
; Sequence 37092, Application US/10369493

Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ. ID NOS: 47374
SEQ. ID NO 37092
LENGTH: 1470
TYPE: DNA
ORGANISM: *Aspergillus nidulans*
US-10-369-493-37092

Query Match 21.8%; Score 317.8; DB 15; Length 1470;
Best Local Similarity 53.8%; Pred. No. 1,3e-83;
Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;

20 TGCCCAACGACTTGAATGGGCTTGCAGCAAGCGCGCTACAGATCGAAGGCGCGTCA 79
2 TGCCCAATGACTTCTTCATGATATGCAAGCGCGCGCGAGTTGAAGAGCGTGA 61
80 AAG 139
62 AAG 121
140 CCAAGCGCGCAAGCGAG 139
122 AAG 161
200 ATCTCTAATGCGGAG 259
182 CCTGATGAG 241
242 TCCCATCGAG 301
320 TGAATGAG 379
302 TGAATGAG 361
380 TGAATGAG 439
362 TGAATGAG 421
440 ACTTGAAG 499
422 ACTTGAAG 481
500 TCAACATCAAG 559
482 TCAACATCAAG 541
560 CGGAG 619
542 CTGACAGATGATGTTTCCGAACTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
620 TCGCTGAG 679
602 TCGTGGGAG 661
680 TCGCTGAG 739
662 TCGCTGAG 721

Qy	1157	GCCCGTGCCTCGAGAGGAAACATGACGTGCGAGGAGACCGGTAAAGACCCCTTTCGCGA	1216
Db	1249	TTGACACAGCCTGGAAATGAGACTTGGCCGTTC-----GCTCTGACACAAATTTCAGGA	1302
Qy	1217	TCGCGTACTTTCGACTCGCACTTGGACTCGATTTTCCAAAGGCCATTACCCAGGACGGCGTCG	1276
Db	1303	TAGACTACTTCGAGAGAGTACCTGTATACGAGCTCCAGTGGCGCATACGC---GACGGTGCAG	1359
Qy	1277	TTCGACAAAGGGGTACTTTTGGCGTGGCGGCGTGTCTCGATTAATTGGAATGTGCABAATGGTACG	1336
Db	1360	ACGCTCTTCGGCTACTTCGCGTGGTGGCGTGGCTGCTGGAACACTTGAATGGGGGCTCGGACTTCA	1419
Qy	1337	GACCCAGATTTCGGGGCTCAGTTCAACAGACTACACCAACCCTGAAGGCGACGGCCAAAGAT	1396
Db	1420	CCTCCAGATTGGGATCGTGTACGTGAGCAGCGGAACAAGTTCGTCGCGGTAACCCCAAGGACT	1479
Qy	1397	CTGGCCTGTGCTCTCAAGGA	1415
Db	1480	CAGCTCGCTGTTCAGGAA	1498

```

RESULT 5
US-10-425-114-19747
Sequence 19747, Application US/10425114
Publication No. US20040034588A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53113)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 19747
LENGTH: 1794
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: L1B3J51-012-C9_F11
US-10-425-114-19747

```

Query Match	15.8%	Score 230.4;	DB 12;	Length 1794;
Best Local Similarity	52.3%	Pred. No. 1.1e-57;		
Matches 722; Conservative	0;	Mismatches 616;	Indels 42;	Gaps 8;

OY	40	AGGCTTCGCAACGGCGCGCTTACAGAGTCGAAGGCGCGTCAAAAGAGGTGGCGCGCGCG	99
Db	237	GGCAACCGCACCTCCTCGTACCAATCGAAGGCGCTACTCTGAGGGCAACAAGACTTG	296
OY	100	TTCATCTTGGACACGTACTGCACTCGAGCCATCGCGACCAACGGCGCCAAACGCGAT	159
Db	297	AGCAACTGGGAGTGTCTTCAGCCACGTAACGAGAAGATGAAGACCGAAGCACCCGAGAT	356
OY	160	GTGGCTTGCGATCAGTACCAACCGCATACGATGAGGACTTGTGATCTCTTGACCAAGTACGCG	219
Db	357	ATCGCGAATGATCACTATCATCTGTTACGAGATGACATTGACTATGCACTCTCTCGCG	416
OY	220	GCAAAAGCGCTTACCGCTTCTCTTGTGTGTGCGGATCATTCGCCCTCGGCGGACGGCTG	279
Db	417	ACCAACGCGCTACAGATTTCTATATCATCGGCGAGAAATCTTCC--GAGAGGGAATTC	473
OY	280	GATCCCGTCAACGAGAGGGAATGAGTTTACAGCAAACTGATTAGCCGCTGTGTAGG	339
Db	474	GGCAGGTCAACCCAGCAGGACATCGGTTCTTCAACAGACTCATGCACTTCGCTCGTCTC	533
OY	340	CGGGGATACCGCTTTGGGTGACTTTGTAACCACTGGGATCTGCTCAGCGGCTTACAGAT	399
Db	534	AAAGGATAGACCGTCTGTGACGCTGTCCCACTAGACATCCCGCAAGAGCTGTGAGAG	593

OY	CGCATATGAGGCTGGCTCAACCTGGAAAGAGTCCAGCTGACCTTTGAGCGGTATGCGAGG	455	
OY	400		
DB	594	AGGTAACGGGGGTGGCTCGCGGGCGCCGAGCGGGCGGCACTTGGCGACCTTGGCGGAC	653
OY	460	TTGTGCTTTGAACTTTTGGGGGACCGAGTCCAGAACTGATATCAACATCAACGAWCCCTGG	519
DB	654	GTGTGCTTTCCGGGGGTTTCGGGGACCGGGTCAAGGCACTGGGCACTTCAACGACCCAAC	713
OY	520	ATTACGCCCATCTATGATATGCCACCGGCAGCAAGCCCGGGGAGAGGACATTAAAC	579
DB	714	GTCCCGGTACACAGGGGCTACATGCTCGGCACTTAACCGCGCGGGGCGCATATG-----TCG	767
OY	580	AAGCACTCCACCGAGGGGCAACATGAGCACTGAGCGGAGGCTCGGTGAAAGGCCAGATC	639
DB	768	CGCTCTTCGCGCCGGGGGCAATCTCGACGCCGAGCCTTAAGTCCGCGGCACACAGTGTCTC	827
OY	640	ATGAGCCATATGCCCGCGCGCGGTGGCGGTCTACAGCAGGAGCTTTCGCCCTCGCAAAAGGC	699
DB	828	CTGGCCCAACCGCGCGCGCTCCAGATCTTACAAG-----CCAAATACCAAGAGGC	878
OY	700	CAGATCCGCATCTTGGCTCAACGGGCACTATATAGCTCTGGGACAGCAATGAGCTTCGG	759
DB	879	TTGATCCGCATCTGATATGTCATCTGTGTGTTCTGTCGCGCTG---ACGGACCGCCCGAG	935
OY	760	GACAAGAGGCTGTGCTGACGACGAGATGGAATTTACATTTGGCTGTTTGCATATCCATC	819
DB	936	GACCGGCTGGGCACTGAGCGGGGCGCTGGCTTCAAGCTTTCATATGTTTCTTGAACCCGATA	995
OY	820	TTCTTGAAGAAGACTATCCAGAGAGCATGAGAAAGAGCTGGCGGAGAGGCTTTCAGACC	879
DB	996	ATCT---ACGGAGCTACCTCCGGAGATGGCGCGGCTCTGGGCTCCAGGCTCCGACC	1052
OY	880	CTCACTCCCGGGACTTTGCCATCTCTCAATGCGGAGAACCGACTTCTACAGGATGAAT	939
DB	1053	TTCTTCGCGGAGATGAGAGAGAGCT---GGGCTACGCGGCTGAGCTTTCATGCGATCAAC	1109
OY	940	TACTACACATCCAGTTTCGC-----GCCGCACTAGACGGTCCCGTCCCGAG	987
DB	1110	CACATCAAGACGCTGTACGCCAGGAGATTGCATGTTCTTCGCCAAGTTATTCGCGTCCGGG	1169
OY	988	ACGGAATACTTCGGCGCCATCATGAGCACCGAGAGATAGAGCGGCGAGCCCCGTTGGC	1047
DB	1170	CAGAGATTCCACCAATCGCTGGCGGCTACACGGGAGAAAGAGCGGAGATTCCCATCGGG	1229
OY	1048	GAGAGAGCGGCTTCGCTGTGCTGCTCTGCGCCCGGACATGTTCCGGAAGCATCTCGCC	1107
DB	1230	CCTCCGACAGGAGATGCCGACGTTCTTACGTGGTTCCTACCGGAGATAGAGAAATGTTCAAC	1289
OY	1108	CGGGTGAACGGCTGTACGGCAA---GCCCATCTACATCAACCGGAACGATGCGCTGC	1164
DB	1290	TACATCAATGAACAATACAGCAATCTCCCATGTTCTTAACCGAAATTTGGCTACGACAA	1349
OY	1165	CCTGAGAGAGAAACATGACCTGCGAGAGGCGCTCAACGACCCCTTCGCTCCGATCCGTAC	1224
DB	1350	GGTGGGAGTGTATTACTACCCGCGGAGAGACTGGCTTGCACGACAGGGAGGATACAG	1409
OY	1225	TTTAACTCGACCTTGAATGATTTCCAAAGCCATTACCAAGACGGGAGTGTGTCAAG	1284
DB	1410	TACCTCGAGGCTACCTCGCAAAACCTGCCAAAGTCACTCAETGACGGGCGCGAATGTCTGT	1469
OY	1285	GGGATCTTTGGTGGGGGTGTGCTGATATCTTGGAAATGTCAGATGGCTACGACCCAGA	1344
DB	1470	GGCTACTTCAATCTGTCTCTCATTTAGCAATTTGAGATGGCTATATGGGTACACCTCGC	1529
OY	1345	TTGGGCGTCAAGTTCAACAGATACCAACCTCAAGGGCACGCCCAAGAAAGTCTGCGCTG	1404
DB	1530	TTGGGCGCTTCACTAGTTGATACCAACGAGAGGAAAGCCAAAGTGTGCGCGCTG	1589

RESULT 6
US-10-425-114-30939
; Sequence 30939, Application US/10425114

```

; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30939
; LENGTH: 1781
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: UC-ZMFLB73106C01_FLI
US-10-425-114-30939

```

Query Match 15.6%; Score 227.4; DB 12; Length 1781;

Best Local Similarity 52.2%; Pred. No. 8,2e-57;

Matches 735; Conservative 1; Mismatches 637; Indels 36; Gaps 9;

```

QY 20 TSCCCACGACTTTGATGAGGCTTCGCAACGCGCCCTACAGATCGAAGGCGCCGCA 79
DB 230 TCCCAAGGCGTTGCGTTCCGGACGCGACCTCGGCTACAGATCGAAGGCGCCGCT 289
QY 80 AAGAAAGTGGCGCGCGCCCTTCATCTGGAGACCTACTGCCACTGAGCCATGCGCA 139
DB 230 CCACCAAGCGCGCGCGCCCTTCATCTGGAGATTCATCGCGACGTCGCAAGAAATATTG 349
QY 140 CCAACGCGCGCAACGCGCGATGCGCTTCGATCACTACACCGCTACAGATAGAGCTTG 199
DB 350 CAGGATATCAAAATGAGACGTTGATGATCACTACAGATGCTACAGAAAGCGTGG 409
QY 200 ATCTCTTACCAAGTACGCGCGCAAGGCTTACCGCTTCTCTTGTGCTGCGGATCA 259
DB 410 ATCTCATGAAAAAGTTGAACTTGTGATGCTACCGGCTTCAATGCTCATGCTCGAGATCT 469
QY 260 TTCCCTCGGCGCGCGCTGATCCCTCAACGAGAGGGAATTGATTTACGCAAC 319
DB 470 TCCCGATGGCGAAGG-----AAATCAATCCAGAAAGGTGACGATTAACATATT 523
QY 320 TGATTCAGCGCCCTGTGAGCGCGGATACGCGCTTGGGATGATTTGATCCAGTGC 379
DB 524 TGATTAACCTATCTGCTTACGAGAGGATGATCTCTTACATCACTTTACATATGATC 583
QY 380 TGCCTCAGCGCTTACGATGCTATGAGAGCTGCGCTCAAGTGAAGAGTCCAGCTGG 439
DB 584 TTCTCTTGGCGCTTACGAGAAATATGAGAGGTGTTAAGCGCAATGCGGACTTG- 642
QY 440 ACTTGAGCGGATGCGAGGTTGCTTGAACGTTTGGGAGCCGAGTCCAGAACTGGA 499
DB 643 --TTTACGACTATGCTGACTTCTGTTTAAAGCTTACGCGGATGCGTAAAGACTGGT 700
QY 500 TCACCATCAACGAMCCTGATTCAGGCACTATGATATGCAACGCGGAGCAACGCC 559
DB 701 TTACATTCATGAGCAAGATAGTAGGCTACTTGGCTATGACAGAGGCTCAATCTTC 760
QY 560 CGGCGAGAGAGCATTTAACAAGACTTCACCGAGGCAACACTGCACTGAGCGCTGGC 619
DB 761 CTCAAAAGGTG-----ACGATGCGCTGCTGAGGAAATTCAGCAACGAACTTACA 814
QY 620 TCGCTGAAAGGCGCAGATCATGAGCCATGCGCGCGCGCTGCTACAGAGGACT 679
DB 815 TAGTTGCTCAATATTTTCTTTGGCAATGCTACTGAGATTGCAAGATCCGTAAGAAAT 874
QY 680 TTGCGCCCTGCAAAAGGCGCATGCGCATCTGCTCAACGCGGACTACTATGAGCCCT 739
DB 875 ATCAGGCTGCTCAGAAAGGTGAAGTCCGAAATAGTCTTGGACTTCACTGATAGAGCTC 934

```

```

QY 740 GGAACGATGAGCTTCGAGACAGAGGCTCTGAGCGACGATGAAATTTCACATTG 799
DB 935 T---TACAACTCAGCTATGACCAAGCAGAGCCCAAGAGCGAGCATTCACATTG 991
QY 800 GCTGTTGGCAATCCCATCTTCTTGAAGAAAGACTTTCAGAGAGATTAAGAGAGC 859
DB 992 GCTGTTGTTGATCC--ATTGATTAACGAGCACTATTCAGATTAATGCAAGATTCG 1048
QY 860 TGGCGAGAGCTTTCAGCCCTCACTCCCGGACTTGGCCATCTTCATGCGGAGAGA 919
DB 1049 TGAAGAGAGGCTCCAGGTTCACTCTAGCAGAGTTAACTGTTAAGGCTCGCAG 1108
QY 920 CCGACTTCTAGGATTAATTC--TACATCTCCAGTTTGGCGCCACTTACAGGCTC 976
DB 1109 ACTACATCGGTATCAAGAGTACATCCAGCTACATGAAAGGCGAAGACTGCTCAGC 1168
QY 977 CCGTCCCGAGAGCGACTATCTCGGCGCCATCCATGAGCAGAGAAATTAAGAGCGCA 1036
DB 1169 TGGGCGCCATGAGTACTCTGCGGATTTGAGAGGTTCAATATGTTTTCAGCGAAATGCA 1228
QY 1037 GCCCGTTGCGAGAGAGAGGCGGCTCGCTGCGCTTCTTCCCGGACATGTTCCGA 1096
DB 1229 AACGATTTGAGACCAAGCGAATTTTAAGTGGCTTACATCGCCCGAGCGGATGTAAG 1288
QY 1097 AGCATCTGCGCGGCTGATAGGCTGTAGCGCAAGC--CACTTCACTACCGAGAACG 1153
DB 1289 GGTGCGTAACTCTTCAAGAGAAATGGAATCAACGATCTCAATTAACGAGAAAG 1348
QY 1154 GATGCCGCTGCGTGAAGAGAGAAATACAGTGCAGAGAGGCGTCAAGACCCCTTCC 1213
DB 1349 GAATGAGCAAGCTGGA-----AATTGACCGGAACAGATGCTGCGGAGCAGCA 1402
QY 1214 GCATCGGTACTTGACTGCGACTTGAATTCATTCAGGCACTTACAGAGCGCG 1273
DB 1403 GGTGCGGTTCTACAGAGCTACATCGGCGAGTGAAGAGGCAATAGACAG--GGAG 1459
QY 1274 TCGTCTCAAGGAGTACTTTCGTTGCGTGGCGTTCGATTAATCTGAAATGCTAGATGCT 1333
DB 1460 CGAAGCTGCTGCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1519
QY 1334 ACGAGCCAGATTCGCGCTACGTTCAAGACTTACCAACCTCAAGCGCAGCCAGAGA 1393
DB 1520 ACTGCTCAAGTTGCGCATGCTTACGTGGACTTCAACGCTCGAAGCCACCGGAGG 1579
QY 1394 AGTCTGCGCTGCTTCAAGAGCATGTT 1422
DB 1580 CGTGCGCTTACTGCTTCAAGGACATGCTT 1608

```

RESULT 7

US-10-425-114-30835

; Sequence 30835, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(5313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 30835

; LENGTH: 1931

; TYPE: DNA

; ORGANISM: Zea mays

; OTHER INFORMATION: Clone ID: UC-ZMFLB73086G03_FLI

US-10-425-114-30835

Query Match 15.6%; Score 227.4; DB 12; Length 1931;
Best Local Similarity 52.2%; Pred. No. 8.4e-57; Indels 36; Gaps 9;
Matches 735; Conservative 1; Mismatches 637;

20 TGCCCAACGATTGAAATGGGCTTCGCAACGCGCCCTACAGATCGAAGGCGCCGTCA 79
236 TCCCAAGGGGTTCTGTTTCGGGACGGGCACTCGGGTACAGAGTGAAGGGCGCCGCT 295
80 AAGAAGTGGCCCGCGCCCTTCCTCTGGGACAGTACCTGCACTGGAGCCATCGGCA 139
296 CCAACACGCGCGCGCCCTTCCTCTGGGATTCATTCGCGCAAGTCCCAAGAAATATG 355
140 CCAACGCGCGCAACGCGGATGGCTTCGATCACTCAACCGCTACGATGAGGACTTGG 199
356 CAGGGAATCAAAATGAGAGCTTTCAGATGATCAATACATCGCTACAGAGAGACGTTG 415
200 ATCTTTCACCAAGTACGAGGCAAAAGGCTACCGCTTCTCTGTCGTGCGGATCA 259
416 ATCTCATGAAAGTTGAACTTTGATGCTTACCGGCTTCTCAATCTCATGCTCCAGATCT 475
260 TTCCCTCCGCGGCGAGCTGATCCGCTCAACGAGAGGAAATGAGTTTACGCAAAAC 319
476 TCCCGGATGGCGAAGG-----AAGTCAATCCAGAAAGGTGAGCGTAATCAATATT 529
320 TGAATGACGCGCTGTTGAGGCGGGGTATCAAGCTTGGGATCTTTGATCACTGAGATC 379
530 TGAATAACTATCTGCTTCAAGAGGCAATGACTCTTCAATCAACCTTTACCATATGATC 589
380 TGCTCAGGCGCTTCAAGATCGCTATGAGAGCTGGCTCAACGTGGAAGAGTCCAGCTGG 439
590 TTCTCTTGCGCTTGAGAAAGAAATATGAGGGTGTTAAGCGGAAATAGGCGGACTTG- 648
440 ACTTTGACGGTATGCGAGTGTGCTTTGAACCTTTGGGAGCGAGTCCAGAACTGGA 499
649 --TTTACAGACTATGCTGACTCTGTTTAAAGCTACGCGGAGTGGGTAAAGCACTGCT 706
500 TCAACATCAACGAGMCCCTGATTCAGGCACTATGATATGCAACGCGGAGCAAGCCGC 559
707 TTAACATTCATGAGCCAGAGATGATGAGCTGCTTGTATGACAGAGGCTCAATCTTC 766
560 CGGCGAGGAGCAGCATTTAAAGCACTCAACGAGGCAACACTGCCATGAGCCGCTGGC 619
767 CTGAAGAGTGC-----ACCAAGATGGCTGCTGGTGGGAATTCAGCAACCGAAGCTTACA 820
620 TCGCTGAAAGGCCCAAGATATGAGCCATGCGCGCGCTGGCGCTTACAGCAGGACT 679
821 TAGTTGCTCAATATTTTCTTGGCACTGCTACTGCAAGTTGCAAGATACGTAAGAAAT 880
680 TTGGCCCTGGCAAAAGGCGCAGATGGGCACTTCGCTCAACGCGGATTAATGAGCCCT 739
881 ATAGAGTGTCTAGAAAGGATGAGTGGAAATAGTCTGGAATTTCACTGATGACAGGCTC 940
740 GGAACAGCAATGAGCTCGGAGCAAGAGGCTGCTGAGCAACGAGTGAATTTACATTTG 799
941 T-----TACAACTCACTGATGACCAAGCAGCAGCCAAAGAGGAGACTTCCACATTTG 997
800 GCTGTTTGGCAATCCCATCTTCTTGAAGAGACTATCCAGAGAGCATGAAGAGCAGC 859
998 GCTGTTTGTGATCC--ATTGATTAACGAGCACTATCCACAGATTAATCAAGATTCG 1054
860 TGGGCGAGAGGCTTTCAGCCCTCACTCGCGGAGCTTGGCATCTCTAAATGCGGAGAGA 919
1055 TGAAGGAGAGGCTGCGCAAGGCTTCTGAGCAGGCTTAACTGGTAAAGGCGCTCGGAG 1114
920 CCGACTTCTACGCGATGAATTTAC---TACACATCCAGATTGCGCGGCACTTACAGCGCTC 976
1115 ACTACATCGGATTCAGAGATGACATCCAGCTACAGTGAAGGCGCAGAACTGCTCAGC 1174
977 CCGTCCCGAGAGCGACTATCTGCGCGCATTCATGAGCAACGAGAGAAATGAAGACGGA 1036
1175 TGGCGCCCACTAGTACTGCGGATTCGAGGTTCAATATGTTTTTGCACGCAATGGCA 1234

1037 GCCCGGTGGCGAGAGAGCGGCTCGCTGCTGCGCTCTGCGGACATGTTCCGGA 1096
1235 AACCGATGGACACACAGGCGAATTTTAAGTGGCTTACATCGGCCCGACGGGATGACG 1294
1097 AGCATCTCGCCCGGCTGTACGCGCTTACGCGCAAGC---CCATCTACATCAGGAGACG 1153
1295 GTGCGTGAATCTACTCAAGAGAGATGAGTGAATCCAGATCTACATTAACGAGAAACG 1354
1154 GATGCCGTCCTTGGAGAGAGAAACATGACGTGCGAGAGGCGCTCAAGACCCCTTCC 1213
1355 GAATGACACAGCTTGA-----AAGTGAACCGAGCCAGTACCTGCGGAGCCACGA 1408
1214 GCATCCGTAATCTTGAATCGCATCTGGAATTCGATTTCAAGGCGCATTTACAGAGCGCG 1273
1409 GGGTGGCGTTCTACAGAGACTTACATCGGCGAGCTGAAGAGGCGCATAGACAG---CGAG 1465
1274 TCGTGTCAAGGGGTATCTTTCGTGGGCGCTTGTCTGATTAATTGAATGTCAGATGCT 1333
1466 CGAAGTGGCTGCTACTTCTGCTGCTGCTCTCGACAACTTCGAGTGGCTGGCAGAGGT 1525
1334 ACGAAGCGAATTCGGGCTTACGCTTACAGACTTACACACCTTCAAGCGGACCCCAAG 1393
1526 ACTGTCGAAGTTTCGAGCATGCTTACGTGATGATCAACAGCTTGAACGCCACCGAAG 1585
1394 AGTCTGCGCTGCTCTCAAGACATGTTT 1422
1586 CTTGCGCTACTGCTTCAAGGACATGCTT 1614

RESULT 8

US-10-425-114-33175
; Sequence 33175; Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yinhua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaka, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 33175
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMPLM017071E02_FLI
US-10-425-114-33175

Query Match 15.6%; Score 227.4; DB 12; Length 1932;
Best Local Similarity 52.2%; Pred. No. 8.4e-57; Indels 36; Gaps 9;
Matches 735; Conservative 1; Mismatches 637;

20 TGCCCAACGATTGAAATGGGCTTCGCAACGCGCCCTACAGATCGAAGGCGCCGTCA 79
237 TCCCAAGGGGTTCTGTTTCGGGACGGGCACTCGGGTACAGAGTGAAGGGCGCGCT 296
80 AAGAAGTGGCCCGCGCCCTTCCTCTGGGACAGTACCTGCACTGGAGCCATCGGCA 139
297 CCAACACGCGCGCGCCCTTCCTCTGGGATTCATTCGCGCAAGTCCCAAGAAATATG 356
140 CCAACGCGCGCAACGCGGATGGCTTCGATCACTACACGCTTACGATGAGGACTTGG 199
357 CAGGGAATCAAAATGAGAGCTTTCAGTGAATTAATACATGCTTCAAGAGAGAGCTG 416
200 ATCTTTCACCAAGTACGCGGCAAGGCTTCTCTGTTGCTGCTGCTGCGGATCA 259


```
Db 417 ATCTCATGAAAAAGTTGAAGCTTTGATGTCCTACCGGTTCTCAATGCTATGCTCCAGGATCT 476
Oy 260 TTCCCTCGGCGGAGGCTGGATCCGTCACAGAGAGGAAATTGATTTTACAGCAAC 319
Db 477 TCCCGATGAGGAGG-----AAAGTCAATCAGAAAGGTATGCGTATTAACAATATT 530
Oy 320 TGATTTAGCGCCCTTTGAGGCGGGGTATCAGCCTTGGGTGACTTTGTACCTAGGATC 379
Db 531 TGATTAACATATCTCTTCAGCAAGGATGACTCTTCATCAACCTTTTACCATATATTC 590
Oy 380 TGCCTCAGGCGCTCAGCATGCTATGAGGCTGAGTCAACGTCGAAAGAGGTCCAGCTGG 439
Db 591 TTCTCTTTCGCTTGAGAAAGAAATATGAGGAGGTGTTAAGCGCAAGATGCGGACTTG- 649
Oy 440 ACTTTGAGCGGTATGCGAGGTGCTTTGAAAGTTTGGGACCGAGTCCAGAACTGGA 499
Db 650 -TTTACAGACTATGCTGACTTCTGTTTAAAGCTACGCGGATGCGCTAAAGACTGGT 707
Oy 500 TCACCATTAAGAMCCCTGAGTTACAGGCATCTATGATATGCGACCGGAGCAACGCGCC 559
Db 708 TTACATTTAATGAGCAAGATAGTAGCTACTTGGCTATGACACAGGGTCAAAATCTTC 767
Oy 560 CGGCGAGAGCAGCATTAACAAGCACTCACCGAGGCAACACTGCCACTGAGCGGTGGC 619
Db 768 CTCAGAGGTGC-----ACAGATGCGCTGCTGGTGGGAATTCAGCAACCGAACCTTACA 821
Oy 620 TCGCTGAAAGGCCAGATCATGAGCCATGCCCGCGCGCTGCTTACAGCAGGAGACT 679
Db 822 TAGTTGCTCATATATTTCTCTTGGGCAATGCTACTGCAAGTTGCAAGATACCGTACGAAT 881
Oy 680 TTCCGCCCTCGCAAAAGGCGCAGATCGCATCTCGTCAAGCGGCACTACTATGAGCCCT 739
Db 882 ATCAGGCTGCTCAGAGGTAAGTGGCAATAGTCTTGACTTCACTGTTGTCAGAGGCTC 941
Oy 740 GGGACAGCAATAGGCTCGGACAGAGGCTGTAGCCGACGAGTGAATTTCCATTTG 799
Db 942 T--TACAACCTCACCTGATGACCAAGCAGCGCCCAAGAGCGAGGACTTCCCATTTG 998
Oy 800 GCTGTTGGCCATCTCCATCTTCTTGAAGAAAGATTAATCCAGAGCACTGAAAGAGCAGC 859
Db 999 GCTGTTGTTGATCC--ATTGATTAACGGAACATATCCACAGATTAATGCAAGATCTCG 1055
Oy 860 TGGGCGAGGCTTCAAGCCCTCACTCCGCGAGACTTTGCCATCTCTCAATGCGGAGAGA 919
Db 1056 TGAAGAGAGGCTGCGCAGGTTCACTCTGAGCAAGGTAATGTTGAAGGCTCGGCGAG 1115
Oy 920 CGACTTTCAGGCGATGAAATTAAC--TACACATCCCAAGTTGCGCGGCCACTTGAAGCGTC 976
Db 1116 ACTACATCGGTATCAACGAGTACACATCCAGCTACATGAAAGGCGAGAGCTGTGTCAGC 1175
Oy 977 CCGTCCCGCAGAGCACTATCTCGGCGCATCCATGAGCACCAGAGAAATAGACGCGCA 1036
Db 1176 TGGCGCCCAAGTACTACTCTGCGATGGCAGGTTCAATATTTTGTGACGCAATGGCA 1235
Oy 1037 GCCCGTTGGCGAGAGAGCGGCTCGCTGCTGCGCTCTGCGCGACATGTTCCGGA 1096
Db 1236 AACCATTTGACACAGCGGAATTTCAATGGCTTCAATCGCCCGAGCGGGAATGTACG 1295
Oy 1097 AGCATTCGCGCGGGGTGTAGCGGCTGTACGCAAGC--CCATTTACATCACCGAGAGC 1153
Db 1296 GGTGCGTGAATTAACCTCAAGAGAAATATGGAATTCAAAGATTAACAATACGAGAAAG 1355
Oy 1154 GATGCCCGGCGCTGAGAGAGAAACATGACGTGCGAGAGGCGCGTCAACGACCCCTTTC 1213
Db 1356 GAATGAGCAGCTGGA-----AACTTACCCGAGACCAATACCTGCGCGAGCGCACGA 1409
Oy 1214 GCATCGGTAATTTGAATCGCACTTGACATTTTCAAAGGCATTTACCGAGAGCGCG 1273
Db 1410 GGGTGGGTTCTTACAGAGGTAATCATGCGCACGCTGAAGAGGCAATAGACAG--GGAG 1466
Oy 1274 TCGTGTCAAGGGGTACTTTGCGTGGCGGTGCTGCATTAATCTTGAAGATGATGCT 1333
Db 1467 CGAAGCTGGCTGCTACTTCTGCTGCTCTCTTCGACAACTTGAGTGGCTGGAGGGT 1526
```

```
Oy 1334 AGGAGCCAGATTGGGCGTACAGGTTCAAGACTACAGCAACCCCTCAAGGCGCAGCCCAAGA 1393
Db 1527 ACTGTTCAGAGTTGGGCATGCTTACGTGAGCTTCAACAGCTTCGAGACCCGCAAG 1586
Oy 1394 AGTGTCCCTGTGCTCTCAAGACATGTTT 1422
Db 1587 GGTGGCGTAACTGTGTTCAAGGACATGCTT 1615
```

```
RESULT 9
US-10-425-114-25892
; Sequence 25892, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 73128
; SEQ. ID NO 25892
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3960-023-C7_FLI
US-10-425-114-25892
```

```
Query Match 15.1%; Score 219.2; DB 12; Length 1737;
Best Local Similarity 51.3%; Pred. No. 2.2e-54;
Matches 727; Conservative 1; Mismatches 654; Indels 36; Gaps 8;
```

```
Oy 20 TGGCCAGCACTTGAATGGGGCTTCCGAACGCGCGCTACAGCATGGAAGCGCGCTCA 79
Db 228 TTCCCAAGGGTGTGTGTGGAGCGCGCGCTGCGCTACAGAGTGAAGGCAATGCGCG 287
Oy 80 AAGAAGTGGCGCGGCGCGCTTCATCTTGGACACGTAATGCGCATGGAAGCATGCGCA 139
Db 288 ACAAGACGCGCGCGCGCGCGCATTTGGGACGCTTCATCAATCAATCCCGCGCAATCG 347
Oy 140 CCAACGCGCCCAAGGCGAGTGGCTTGGCATCTACCAACCGCTACAGTGAAGGACTTTG 199
Db 348 CAACCAACGCGCGCGGACGTGACTGTGACGATCCATCGCTTACAGAGGACGTGA 407
Oy 200 ATCTTTGACCAAGTACGCGCGCAAGGCTTACCGCTTCTCTGTGTGTGCGGATCA 259
Db 408 ATATCATGAAATAATATGGGGTTGACCGCTTACCGGTTTTCATCTTGTGTCAAGAAATAT 467
Oy 260 TTCCCTCGCGCGGAGCTGATCCGTCACAGAGAGGAATGAGTTTACGCAAC 319
Db 468 TCCCAATGGAAC-----GAGAGTGAATGGAAGAGTGGCGTACTCAACAGGCG 521
Oy 320 TGATTTAGCGCTGTGTTGAGCGGGGTATACAGCTTGGGTGACTTTGTACCATGAGATC 379
Db 522 TGATTAATTAATGTGAAGAAAGGATCACACCTACGCAACCTGTACCACTACAGACC 581
Oy 380 TGCCTCAGCGCTTACAGATGCTATGAGGCTGAGCTCAACGTGGAAGAGGTCCAGCTGG 439
Db 582 TACCGAGAGCGCTGAGGTTCGGTACGAGAGGCTGTTGAGAGAGAGTGTG---GAGAT 638
Oy 440 ACTTTGAGCGGTATGCGAGGTGTTGCTTTGAACGTTTGGGGAACGAGTCCAGAACTGA 499
Db 639 CGTTGCGAGACTACGCGCACTTGTCTTGGGGGCTTGGGAGCAAGGTGAAGACTGGC 698
Oy 500 TCACCATCAAGAMCCCTGATTCAGGCACTTATGATATGCAACCGGAGCAACGCGCC 559
```

```

Db      699 TGAAGTTCACGAGCCGCGGGGTGGTGGCCGCTTAGGGTAAGAGCGAGGTTCCGCG 758
Qy      560 CGGCGAGAGACAGCATTAACAGCACTCAACGAGGGCAACTGCACTGAGCCGCGG 619
Db      759 CGGAGAGGTTCAGGGGTGGAG-----GCCGGGGGGGACTGGGCAACGAGGCTTAAG 812
Qy      620 TCGCTGAAAAGGCCAGATCATGAGCCATCCCGCGCGCTGGCGCTTACAGAGGAGACT 679
Db      813 TCGTGGGAGCAACCTCATCTCTCCAGCGCCGCGCGCTGCAAGAGTACCGCGCAGGC 872
Qy      680 TTGCGCCCTTCGCAAAAGGGCAAGTGGCATCTGCTCAACGGGCACTATGAGCCCT 739
Db      873 ACCAGCCGACGCGAGAGGGGCGAGGTCTGCTGAGATTCTGTGTGTACGAGGCC 932
Qy      740 GGGACAGCAATGAGCTCTGGGACAAGAGGCTGTGAGCGACGAGTGAATTCACATG 799
Db      933 TACGGCGGAGCTAGCGCGGACCGGGCCCGCTCAAAAGTCCAGAGACTTCACACTCG 992
Qy      800 GCTGTTTTCGCAATCCCATCTTCTTGAAGAAGACTATCCAGAGAGCATGAAGAGCAGC 859
Db      993 GATGTTCTCGACCCCATCTGCT---ACGGCGAGTACCCCAAGTGGTCCGGAGAAAGC 1049
Qy      860 TGGGCGAGAGGCTTCAAGCCCTCACTCCGCGGACTTTGCTCATCTCATATGCCGAGAGA 919
Db      1050 TCAAGGGCAGGCTCCCAAGTTCACGCGTGAAGG---GGCGGCTTAGTCCGAGGCTCCA 1106
Qy      920 CGGACTTTCAGGCGATGAATTAACAATCCAGTTCCGCGCGCACTAGAGAGGTCGCG 979
Db      1107 TCAGACTAGTGGAGTCAACAGTACCTGCTTACGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1166
Qy      980 TCCCGAGAGCG-----GACTATCTCGGCGCCATTCATGAGCAACGAGAAATTAAG 1030
Db      1167 CTAGGGGGGGGGCGCGCCAGCTACTCTGCTGCACTGGCACTGGAGTTGCTATGAAAGCG 1226
Qy      1031 ACCGAGAGCCCGTTGGCGAGAGAGCGGCTCTGCTGGCTGCGCTCTGCTGCGGACATGT 1090
Db      1227 ACCGTTGCGGATTTGACCAAGGCGCAACTCAGACTGGCTCTACATGTCCTTTGGGGAC 1286
Qy      1091 TCCGGAAGCATCTGCGCGCGGGGTGAAGCGCTGTACGGCAAGCCCATTAATACGAGAGA 1150
Db      1287 TGTACAAAGCCGTACCTTACGTCAAGAGAGTACGGCAACCCCA---CGATCTCTGT 1343
Qy      1151 ACGGATGCCGCTGCGCTGAGAGAGAACATGACGTGCGAGAGCGCCGTCAAGACCCCT 1210
Db      1344 CAGAGAACGGTATGAGAGAGACCGGGCAACCTCACGTTGGGCGAGGGGTCCGAGACGG 1403
Qy      1211 TCCGCAATCCGATCTTGAATCTGCACTTGAATTCAGAGCCATTAACCAAGAGAG 1270
Db      1404 CGAGGGTGGCTACT---ACCGAGGCTACGTCGCCGAGCTGAAGGCGGCGATCGAGCG 1460
Qy      1271 GCGTCTGCTGCAAGGGGTACTTTGCGTGGCGCTTGTCTGCTGAATTAATTGGAATGTCAGATG 1330
Db      1461 GCGCAACTGCGTGGAGTACTTGGCTGCTGCTGCTGCTGCACTTGAATGGAAGTGG 1520
Qy      1331 GCTACGAGACCAATTCGCGGTCAAGCTTCAAGACTACCAACCTCAAGACGCGCA 1390
Db      1521 GGTACACGTCGCGGTTCGCGCTCTCTCTACGCTCACTTCAAGAGGCTCCGCGCTACCCCA 1580
Qy      1391 AGAAGTCTGCGCTGCTCTCAAGAGCAATGTTTGGCGGC 1428
Db      1581 AGAGCTCGCGGTACTGTTCAAGGATGTCAATCGCCGCG 1618

```

```

RESULT 10
US-10-425-114-13710
; Sequence 13710, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E

```

```

; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 13710
; LENGTH: 1759
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB143-030-G7_FLI
US-10-425-114-13710

```

```

Query Match      12.9%; Score 187.6; DB 12; Length 1759;
Best Local Similarity 51.0%; Pred. No. 5e-45;
Matches 663; Conservative 1; Mismatches 600; Indels 36; Gaps 8;

```

```

Qy      138 CACCAACGGGCGCAAGCGATGCTGGCTTGCATCACTACACCGCTAGATGAGACTT 197
Db      368 CGCAACCAAGCCACCGCGAGCTGACTGTGACGATACATCGCTTCAAGAGAGACT 427
Qy      198 TGATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTGTGTGTGCGGAT 257
Db      428 GAATATCATGAAGATATGCGGTTTCAGCGCTTACCGGTTTTCGATCTTGTGCAAGAT 487
Qy      258 CATTCCCTTCGCGGCGAGCTGATCCCTCAACGAGAGGAAATGAGTTTACAGCA 317
Db      488 ATTCACCAATGGAACCGAGAA-----GTGAATGGAAGAGTGGCGTACTACACAG 541
Qy      318 ACTGATGAGCGCTGTTGAGGCGGGGTATACAGCCCTGGGTGACTTTGTACACTGGA 377
Db      542 GCTGATTAATCATGTGTAAGAAAGCATCACCTTACGCAACCTGATCAACATCA 601
Qy      378 TCTGCTCAGGCGCTTACGATGCTATGAGGCTGCTCAAGGTGAAGAGGTCCAGCT 437
Db      602 CTACCGGAGGCGCTGAGGTCGCTACGAGGCGTGTGAGCAGAGAAAGTGT---GAG 658
Qy      438 GGAATTTGAGCGGTATGCGAGTTGCTTGAACGTTTGGGAGCCGATCCAGAACTG 497
Db      659 ATGCTTCGAGATACGCGCATCTTCTGCTTGGGGGCTTGGGAGACAGGTGAAGACTG 718
Qy      498 GATCACATCAAGAACCTGATTCAGGCGCATCTATGATATGATGCAACCGCGCAAGCG 557
Db      719 GCTGAGCTTCAAGAGCGCGGGGTGTGCGGCTTACGAGTACAGAGCGAGTTCCG 778
Qy      558 CCGGCGAGGAGCAGATTAACAAGACTTCAACGAGGCAACCTGCACTGAGCCGTG 617
Db      779 GCGGGGAGGTGCACGGGGTGCGAGG-----CGGGGGGGAATCGGGGACCGAGCCCTA 832
Qy      618 GCTGCTGGAAGAGCCCAATCATGAGCCATGCGCGCGCGCTGAGCCGTCTACAGCAGGA 677
Db      833 CGTCTGTGGCGCACACACTATCTTCCACGCGCGCGCTGTCAGAGGATACCGCGCAG 892
Qy      678 CTTTGGCCCTCGCAAAAGGGCCAGATTCGCGCATCTGCTCAAGGGGACTACTATAGCC 737
Db      893 GCAACAGCGCAACGAGAGGGGAGGATCGGATCTGCTGATTTGTTGTGTAAGAGCC 952
Qy      738 CTGGGACAGCAATGAGCTTCGAGCAAGAGAGGTGTCAGCGACGAGTGAATTTTCAT 797
Db      953 CTTCAAGGGGAGTCAAGCGCGGACCGGGCGCGCTCAAAAGTCCAGAGACTTTCACGT 1012
Qy      798 TGGCTGTTTTCGCAATCCCATCTTCTTGAAGAAAGACTATCCAGAGAGCATGAAGAAGCA 857
Db      1013 CGGATGTTCTGCAACCCCATCTCT---ACGCGAGTACCCCAAGTCCGCTCCGAGAG 1069
Qy      858 GCTGGCGAGAGGCTTCAAGCCCTCACTCCCGGAGCTTTGCACTCTCAATGCGGAGAG 917
Db      1070 CGTCAAGGGGAGGCTCCCAAGTTCAAGGCTGAGG---GCGCGTCTAGTCCGAGGCTC 1126
Qy      918 GACGACTTTCAGGAGTGAATTAACAATCCAGTTTGGCGGCGACCTAGACGCTCC 977

```

Db	1127	CATGCACTACGTGGAGTCAACAGTACACTGCTACTACGTGTGTATGACGGCCAA	1186
QY	978	CGTCCCGAGACG-----GACTATCTGTGGCCATCCATAGACAACAGGAATTA	1028
Db	1187	CGCTACGGCGGGCGCGGCCAGCTCTCGTCCGACTGTGCACGCTGATGTTCTATGAAACG	1246
QY	1029	GGAAGGACGACCCCGCTTGGCGAGAGAGGGCGCTGCGCTGGGTGGCTCTCCCGGACAT	1088
Db	1247	CGACGGGTGTGCCAATTGGACCAAGGGGGAATCTAGACTGTGCTTACATCTGTGCTTGGGG	1306
QY	1089	GTTCCGGAGACATCTCGCCGGGGTGTACGGGCTGTACGGCAAGCCCATTTACATCCGA	1148
Db	1307	ACTGTACAAAGCCGTCACACTACGTCAAGAGAAAGTACGGCAACCCCAAGATCTCTGTC	1366
QY	1149	GAAAGGATGCCCGTGTCCCTGGAGAGAGAAATGATGACGTGGCGAAGGCGCGTCAACGACC	1208
Db	1367	AGAGAACGATATGAC--GACCGGGCAAGCTCAGGTGGGCCAAGGGCGTGTGGCGACGC	1423
QY	1209	CTTCGCACTCCGCTACTTTGACTCGCATCTGAACTCGATTTCAAGGCGCAATTACCAGA	1268
Db	1424	GGCGAGGGTGGCTACT--ACGGAGCTACGTGCGCGAGGTGAAGGCGGCGATGACGG	1480
QY	1269	CGGCGTGGTGTCAAGGGGTAATTTGCGTGGCGGTTGTGCATATACTTGAATGGTGA	1328
Db	1481	CGGCGCCAACTGCGTGGAATCTTGCTGCGTGTGCTCGACAACTTCGAGTGAAGCT	1542
QY	1389	TGGCTACGGACCCCAATTCGCGGTGTCAGCTTCAACAGACTAACCACTTCAAGCGACGCC	1388
Db	1541	GGGGTACAGTCCCGGTTCCGGCTCTGTCTACGTGCACTTCAGAGACGCTCCGGCGTACCC	1600
QY	1389	CAGAAGTCTGCGCTGTGTCCTCAAGAACATGTTTGGCGCC	1428
Db	1601	CAGAAGCTCGGGCTACGTGTTAGGGAATGATCGCGGC	1640

```

RESULT 11
US-10-425-114-24850
; Sequence 24850, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCES: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 24850
; LENGTH: 1735
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LB3689-207-B2_F11
US-10-425-114-24850

Query Match          12.7%; Score 184.4; DB 12; Length 1735;
Best Local Similarity 50.3%; Pred. No. 4,4e-44;
Matches 710; Conservative 0; Mismatches 666; Indels 36; Gaps 9;

QY      16 GCCTCCGCCAAGACTTTGAATGAGGGCTTCGCAACGCGCGCTCAACAGATCGAAGCGCC 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      166 GCGTTCGCCGACGCGCTTCGCTTCGCGACGCGCGCGCTCGGCGTACCAAGTCCAGGGGATG 225
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      76 GTCAAGAAAGTGGCGCGCGCGCGCGCTCCATCTGGGACACGTACTGCCACTCGAGCCATCG 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      226 GCCAAGACGCGGGCGGGCGGGGCCCCAGCATCTGGGACCGCTTCAATAGAGTTCCTGGGACC 285
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      136 CGCACCAACGCGCCCAACGCGCATGTGTGGCTTGGCATCATCAACACGCTCAAGTAGAGAC 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	286	ATCCCTAACAAATCCACCGCTGACGTCAGTCGTCAGCATATCATCGTATCAAGAAAGAT	345
Oy	196	TTTGATATCTTTGACCAAGTACGGCGCAAGGCGCTACCGCTTCTCTTGTGTGTGCGG	255
Db	346	CTGAACATATGAAAGAACATGGGCTTTGATGCGTACCGGTTTTCCATCTCTTGTGTGAG	405
Oy	256	ATCATTTCCCTCGCGGACAGGCTGTGATCCCGTCAAGAGAGGGAATTGATTTTACAGC	315
Db	406	ATTTTCCCAATGAAACTGGCAG-----GTGAACGAGAAAGATGTGATTACTTCAAC	459
Oy	316	AACTGATTTGACCGCCTGTGTGAGGCGGGGTATCAGCCCTTGGGTGACTTTGTACACTG	375
Db	460	AGGCTCATATATTAATCATATGCTCCAGCAAGGATATGGCCCGGTATCAATCTTACCATAT	519
Oy	376	GATCTGCTCAGGCGCTTCAAGATCCGTATGAGAGCTGTGCTCAACGTGAAGAGGTCAG	435
Db	520	GACCTCCCATTTGGACATCCATGAACAGTACTGGGGCTGGCTTAGC---CCAAAGATTGTG	576
Oy	436	CTGACCTTTGAGGGGTATGCGAGGTTGTGTCTTTGAAAGTTTGGGAGCCGAGTCCAAAC	495
Db	577	GAGCGCTTTGACACTACGCCGAGTTCTGTTCCACGCGTTGGAGACAGGGTGAAGAAC	636
Oy	496	TGATTCACATCAACGAMCCCTGATTCAGGCCATCTATGATATATGCCACCGGACCAAC	555
Db	637	TGTTTACCTTCAACGACCGAGGTGCGTGTGCTCTGTGGCTTACACAAATGTGCTTGAC	696
Oy	556	GCCCCGGGACGAGCAACAATTAAACAAGCACTCCACGAGGGCAACATGCACTGACCGG	615
Db	697	GCACCGGGAAGGT-----TTCCGGGTGCCCGCCGAGAGCAACTCAACACGAGGCGG	750
Oy	616	TGGCTCGCTGMAAGGCCAAGATCATGAGCCATGCCCCGCCGTTGCCGTTCACAGCAGG	675
Db	751	TACCTGTGCGACACCATCTCATCTTCTTCATGACAGCTCCGGTCAAGCGCATCCGGAC	810
Oy	676	GACTTTGCCCCCTCGCAAAAGGCGCAATGGGATCTCCGCTCAACGCGCACTACTATGAG	735
Db	811	AAAGTACAGCTTCAACGAAGGGAGAAATTGGAATTTCTCTGATTTTCGTGTGTACGAA	870
Oy	736	CCCTGGGACAGCAATGAGCTTCGGGACAAAGAGGCTGTGAGCGACGATGGAATTTTAC	795
Db	871	CCTTT---CAGCGACAGCAATGCGGACCAAGGCTGCACACACAGCGAGCAGGAGACTTCAC	927
Oy	796	ATTGGCTGTTTGGCAATCCATCTTCTTGAAGAAGACTATCCAGAGAGCATGAAG	855
Db	928	CTAGGCTGTTCCTTGACCCCA---TTGTACATGAGCGGTACCCGTACTGTGATTCAGAG	984
Oy	856	CAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGACCTTTGCACTCTCAATGCGGA	915
Db	985	ATTGCGAAAGACAGGCTACCGTTGTTCAGGATGAAGAGCAGAGTGTGAAAGGCTCT	1044
Oy	916	GAGACCGACTTCAAGCGATGAATTACTACACATCCAGTTCCGGCGGCACCTAGAGGCT	975
Db	1045	ATGACCTATGTTGGATCAACACACTACACTTCTTTCATCAAGAAAGACCTCGGACATGG	1104
Oy	976	CCCGTCCCGAGACGGACTATCTCGGCGCCATCCATGAGCAACAGAGAAATAG---GAC	1032
Db	1105	AACTGAGCCAGTACGTATCCAGGATATTTGGATATTTGTCTAGAACGAAT	1164
Oy	1033	GGCAGCCCGTTGGGAGAGAGCGGCTCGCTGGCTGGCTCCTGCGCGGACATGTTT	1092
Db	1165	GGAGTTCTTAATGGGCTCAAGCAAACTCTACTGCGGTCAATTTGTCGCGGCGATC	1224
Oy	1093	CGGAAGACTTCGCGCGGGGTGAAGCGCGCTGATACGGCAGC---CCATCTAATGACCGAG	1149
Db	1225	AACAAGGCTGATGCTATGTCAAGAAATTCAAAAATCTTCAACATGATCTTGTGAA	1284
Oy	1150	AAAGGATGCGCGTGCCTGAGAGAGAACATGACGTGCGAGAGCGGCTCAACGACCCC	1208
Db	1285	AAGGAAATGACCAACCTGTGATGTCAGTATTAAC-----TACGGGTGTGATGACACA	1338
Oy	1210	TTCCGATCCGATCTTTGACTTCGCACTTGGACTGGAATTTCCAAAGCAATTACCGAGAC	1265

Db 1339 GTAAGATCCGTTA---TTACAGAGACTACATTAAGCTCAAGAAAGCAATAGATGAT 1395
Qy 1270 GGGGTGCTGCTCAAGGGGTACTTTGGGTGGCGTTGGCTGATACTTGAATGGTAGAT 1329
Db 1396 GGTGCGAGAGTCTTTGGGTACTTTGGGTGGCTGGCTGCTTGAACAATTGAGTGAAGCTT 1455
Qy 1330 GGCTACGAGACCGAGATTCGGCTGACGTTTCAACAGACTACCAACCTTCAAGCCGACGCC 1389
Db 1456 GGGTACACTTGGGGTTGGCTTGGCTTGGTACGTGAGACTACAAAGACTGGAAGAGTACCC 1515
Qy 1390 AAGAAGTTCGCTGCTGCTTCAAGAGCATGTT 1421
Db 1516 AAGGACTCAGCTTTCGTGTTCAAGCATATGCT 1547

RESULT 12
US-10-425-114-27741
; Sequence 27741, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 27741
; LENGTH: 1757
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4739-015-E1_FLI
US-10-425-114-27741

Query Match 12.7%; Score 184.4; DB 12; Length 1757;
Best Local Similarity 50.3%; Pred. No. 4,4e-44;
Matches 710; Conservative 0; Mismatches 666; Indels 36; Gaps 9;
Qy 16 GCTCTGCCCAAGACTTTGAATGGGCTTGCACAAGCCGCCCTCAACGATGGAAGGCC 75
Db 188 GCTTCCCGGAGCGCTTCTTCTGAGCGCGCGCTGCGTCAAGGTCGAGGGATG 247
Qy 76 GCTAAGAAGAGTGGCGCGCGCGCTCCATCTGGGACAGTACTGCCACTGAGCCATCG 135
Db 248 GCCAAGCAGCGCGCGCGCGCGCGCGCTTGGAGCCCTTCATAGAGTTCCCTGGGACC 307
Qy 136 CGCAGCAACGGCGCCCAAGCGGATGCTGGCTGGCATCACTACCAACGCTGACATGAGAC 195
Db 308 ATCCCTTAACATGCTCAACCGCTGACGTGACGATGACGATGATCATCGTCAAGAAATG 367
Qy 196 TTGTGATCTCTTGAACAAGTACGGCGCAAAAGCCCTACCGCTTCTTGTGCTGTCCGG 255
Db 368 GTCAACATAATGAAGAACATGGCTTTGATGCGTACCGGTTTTCATCTCTTGGTCAGG 427
Qy 256 ATCATTCCTCCCTCGCGCGGAGCTGATCCGCTCAACAGAGAGGAATTGATTTACAGC 315
Db 428 ATTTTCCCATGAGTACCTGGCAG-----GTGAACAGAGAGAGAGTGTACTACAC 481
Qy 316 AAACTGATGACCGCTTGTGAGCGGGGATACACGCTTGGGTGACTTTGTACACTGG 375
Db 482 AGGCTCATAGATTAACATGCTCCAGCAAGATTCGCGCGTATGCAAACTCTACCATTA 541
Qy 376 GATTCGCTCAGCGCTTCAAGATCGCTATGAGGCTGGCTCAACGTGAAAGAGTCCAG 435
Db 542 GACCTCCCATTTGGATCCCATGAACAGTACCTGGCTGGCTTGGC---CCAAAGATTGTG 598
Qy 436 CTGAGCTTTAGCGGTATGCGAGGTTGTGCTTGAACGTTTGGGAGCCAGTCCAGAC 495

Db 599 GAGGCGTTTGACAGACTACCGCCGAGTTCTGCTTCCAGCGCTTCGAGACAGGGTGAAGAC 658
Qy 496 TGGATACCATCAACGAMCCCTGAGATTCAAGGCATATGAAATATCCACCGGACAGAC 555
Db 659 TGGTTTACCTTCAACGAGCGAGGTGCTGTGCTGTGGCTTACACAAATGACTTGCAC 718
Qy 556 GCCCGGAGAGAGAGCACTTAACAAGCACTCCACCGAGGCAACACTGCACTGAGCCG 615
Db 719 GCACCGGAAAGTG-----TTCCGGGTGCCCCCGGAGGCACTCCACAGAGCCG 772
Qy 616 TGGCTGCTGAAAAGGCCAGATCATAGCATGCCCCGCGGTGACCTTACAGCAGG 675
Db 773 TACCTTGCGACACCATCTCATCTTCTCACTGACAGTGCAGTCCGCTCAGGCGATCCCGAC 832
Qy 676 GACTTTCGCCCTCCGAAAGGCCAGATCGCATCTCCCTCAACCGGCACTACTAGAG 725
Db 833 AAGTATCAGCTTCCACGAGAGGGAAAGATTGAAATTCCTGATTTCTGTGTGACGAA 892
Qy 736 CCTGGGACAGCAATGAGCTCGGACAAAGAGCTGTGAGCGACGATGGAATTTAC 795
Db 893 CTTTT---CAGCGACAGCAATGCGGACGAGCTGACAGCACAGGACCGAGACTTCCAC 949
Qy 796 ATTGCTGTGTTGCAATCCCATCTTCTGAAGAGACTATTCAGAGCATGAAGAG 855
Db 950 CTAGGCTGTGTTCTTGACCCCA---TTGTACATGACGAGTACCGTACTGATGCAAGAG 1006
Qy 856 CAGCTGGGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTCCTCATATGCCGA 915
Db 1007 ATTGCCAAGACAGGCTTACCGTGTGTTTACAGATGAAGAAAGCCAGATGTGGAAAGCTCT 1066
Qy 916 GAGACGACTTCTACGAGCAATGAATTAACAATCCAGTTTCGCGGCCACTAGACGT 975
Db 1067 ATGAGCTATGTTGGATCAACCACTACACTCTTCTTCAATGAAGAACCTGGGACATGG 1126
Qy 976 CCGTCCCGGAGACGAGCTATCTCGGCGCCATCATGAGCACAGAGAAATAG---GAC 1032
Db 1127 AACCTGACGCACTACAGTACAGATGATGAGTATGTCATGTTTGTCTACGAACGAAT 1186
Qy 1033 GCGAGCCCGCTTGGCGAGAGAGCGGCTCGCTGTGCTGCTCTCCCGGACATGTTTC 1092
Db 1187 GAGATTCATATGCGCTCAACGCAAACTCTTCTGTGCTGATCATTTGCTGTGGGCAATC 1246
Qy 1093 CGGAAGCATCTCGCCGCGGTGATCGGCTGTACGCGAAGC---CCATCTTACATCACGAG 1149
Db 1247 AACAGAGCTGTCACTATGTCAAGAAACCTTACAAAAATCTTCAATGATCTTGTGAA 1306
Qy 1150 AACGATGCGCGTCCGTGAGAGAGAAACATGACGTGCGAGAGGCCGTCAAAGACCCC 1209
Db 1307 AACGAAATGAGACCAACCTGATGATGCAATTAAC-----TCAGGAGTGTGATGACACA 1360
Qy 1210 TTCCGATCCGATCTTGAATCTGCACTGAGATCGATTCGATTTCCAAAGCCATTAACAGAG 1269
Db 1361 GTAAGATCCGTTA---TTACAGAGACTACATTAATGAGCTCAAGAAAGCAATAGATG 1417
Qy 1270 GGGCTCGCTCAAGGGGATCTTGGGTGGCGCTTGTGCTGATTAATTTGATGAT 1329
Db 1418 GGTGCGAGAGTCAATGGGTACTTTGGGTGTGCTGCTGCTTGAACAATTGAGTGAAGCTT 1477
Qy 1330 GGCTACGAGACCGAGATTCGGCGTCAAGTTCAACAGACTACCAACCTTCAAGCGCACGCC 1389
Db 1478 GGGTACACTTCCGGGTTTGGCTTGTGCTGATGAGTACCAAGACTGTAAGAGAGTACCC 1537
Qy 1390 AAGAAGTTCGCTGCTGCTTCAAGAGCATGTT 1421
Db 1538 AAGGACTCAGCTTTCGTGTTCAAGCATATGCT 1569

RESULT 13
US-10-369-493-31938
; Sequence 31938, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIORITY FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIORITY FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 31938
LENGTH: 1338
TYPE: DNA
ORGANISM: Thermobifida fusca
US-10-369-493-31938

Query Match 12.6%; Score 183.2; DB 15; Length 1338;
Best Local Similarity 51.7%; Pred. No. 9.3e-44;
Matches 714; Conservative 1; Mismatches 589; Indels 78; Gaps 10;

QY 31 TTTGAAATGGGGCTTGGCAACGGCCGCTACCAAGATCGAAGGCGCGCTCAAGAAAGTGGC 90
DB 1 TTTGATGGGAGAGTGGCAACCGCTTCTTCCAGATCGAAGGCTCCACACGCGCGAGCGC 60
QY 91 CGCGGCGCGCTCATCTGGGACACGTAAGTCCGCACTGGAGCATCGGCAACGCGCGC 150
DB 61 CGCGGCGCGCATCTGGGACACCTTTCGCGCACTCGCGGCAAGTGAAGAACGCGCAC 120
QY 151 AACGCGATGTGGCTTGGATGATCAACCGGCTACGATGAGAGATTGATCTTGGACC 210
DB 121 ACGGCGACCTGCTGGGACCACTAACCGGTAACCGGATGAGGTGCTTGAATGGG 180
QY 211 AAGTACGCGCAAGAGGCTACCGCTTCTTCTGTGTGTGGTGGGATATTCCTCGGC 270
DB 181 GAGCTGGGCGGTGGGCGCTTTCATCGCTGGCGCGGATCGAGCCGAGGAGC 240
QY 271 GCGAGGCTGGATCCGCTCAACGAGAGGAGATTGATTTCAGCAAACTGATTGACCC 330
DB 241 -----AAGGGCACCGCGCTGGAGGCGGGCTGAGACTTTCAGACCGGCTGTGACTGC 294
QY 331 CTGTGAGCGGGGATATCAAGCTTGGGATCTTGTATCACTTGGGATCTGCTGAGCG 390
DB 295 CTGCTGAGGCGGCAATCGAGCGGTGGCCGAGCCCTTACACTTGGGACTGGCGAGCG 354
QY 391 CTTCAAGATCGCTATGAGGCTGGCTCAACGTGAGAGAGTCCAGCTGACTTTGAGCGG 450
DB 355 CTGGAGGA-----CGCGGCGGCTGGCCGACCGGACACGCGCAAGCGG---TTCCCGCAGC 408
QY 451 TATGAGAGTTGTGCTTTTAAAGCTTTTGGGAGCCGAGTCCAAATCTGTATCAACATCAAC 510
DB 409 TACGCGGAATCGTCTTACCGCGGCTCGGCGAGCGGATCAACCACTGGAACACGCTCAAC 468
QY 511 GAWCCCTGATTCAGGCACTTATGATATGCAACCGGAGAGACGCCCGGCGAGAGC 570
DB 469 GAGCCGTGTGCTCCGCGTTCCTGGGCTAGGCTCCGGGCTGACCGCCCGGGCC----- 523
QY 571 AGCATTTACAGACATCAACGAGGCAACACTGCACTGAGCGGTGCTCGCTGGAAG 630
DB 524 -----GCCAGAGCCGGCTGCTGCGCTGGCGCGCGCCGAC 558
QY 631 GCCCAATCATGAGCCATGCGCGCGCGGCTGGCGCTTACAGAGGAGCTTTGGCCCTG 690
DB 559 CACTTATGTGTGGCCACGCGGCTGGCGCGCTGGT-----ATGCGGAGCTTTGGCGGCGAG 615
QY 691 CAAAAGGCGCAGATCGCATCTCGCTCAACGCGGAGCTACTATGAGCCCTGGGACAGCAAT 750
DB 616 GCGGAGCTTCCGTGGGATCGGTGTGGGCAACAACAAGCAGGATCGGTCTCTACACT 675
QY 751 GAGCTTCGGGACAGAGGCTGCTGAGCGAGGATGGAATTTCACTTGGCTGTGGC 810

DB 676 GACAGTAGGCGCGACCGGAGCGCTGCGCGCGGATTTGACGCGCTGCGGAACCGCATCTTC 735
QY 811 AATCCATCTTTCTTGAAGAAGACTATCCAGAGAGATGAAGAAGAGCTGGGCG----- 865
DB 736 ACCGAGCCGCTGTGAAGGCGCGCTTACCGGAGACCTGATGAGAGACGTCCCGCGGTC 795
QY 866 -----AGAGGCTTTCAGGCTTCACTCCGCGA-----CTTGGCATCTCAATGCG 912
DB 796 ACCGACTACAGCTGTGCTCCAGAGCGGCGACCTGAAGACCATCTCCGCAACCTGAGCATG 855
QY 913 GGAAGACCGACTTCTTACGCGATGATTTACTACATCCCATTTGGCGCGCACTAGAC 972
DB 856 ATGGGCTCAACTTCTTCAACCGGAGCTGAGTGGTGTGAGCAACCGGAGAACCGGGCTCC 915
QY 973 GATCCGCTCCCGGAGACGAGACTATCTCGGCGCATCATAGGACCCAGAGATTAAGAC 1032
DB 916 GACCGGCTGCGGACGAGGCTTACTGCGCGTCTG-----GGCAGGAGCATGTCTGGAG 972
QY 1033 GCGAGCCCGCTTGGCGAGAGAGAGCGGCTGCGCTGCGCTCTGCGCGCATGTTTC 1092
DB 973 GTGGACCCCGGCTGCGCGGTGACCGGCATGGGCTGGCGGATGACCCGACCGGCGGTAC 1032
QY 1093 CGGAGCATCTGCGCGGCTGTACGCGCTGTACGCGAGCGCATCTTACATCACCGAGAC 1152
DB 1033 GACACGCTACCGGCTGGCAACGACTACCGGAGCTGCGCTGTACATCACCGAGAAC 1092
QY 1153 GATGCGCCGCTGCGAGAGAGAAATGACCTGCGAGAGAGCGCTCAACGACCCCTTC 1212
DB 1093 GCGCGCGCTTC-----GAGGACAAAGTGTGTGACGCGCGGTGACGACACCGAG 1143
QY 1213 CGATCCGTAATCTTGAATCTGCACTTGGACTGATTTCCAGGCGCTTACCGAGAGCG 1272
DB 1144 CGATCGCTTACCTGAGCTGCACTGCGGCGCGGCGGCAAGCTGCACTT-----GAGGCGGCG 1200
QY 1273 GTGCTGCTCAAGGGGTACTTGTGGGCGTTCCTGATCTGATCTTGAATGTGATGCG 1332
DB 1201 GTGCGCTCAAGGGGTACTTGTGGCTGTGCTGTGCTTATGAGCACTTCAAGTGGGCGCTCGGG 1260
QY 1333 TACGAGCCAGATTCGCGCTCAAGCTTCAACGACTACCAACCTTCAAGCGGCGGCAAG 1392
DB 1261 TACGGAAGGGGTTCGCGCATGTGCACTGAGCTGAGCTACGAGAGCGGCGGAGTGAAG 1320
QY 1393 AA 1394
DB 1321 GA 1322

RESULT 14
US-09-938-842A-1282
Sequence 1282, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Krepes, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
PRIORITY FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1282
LENGTH: 1734
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1282

Query Match 12.6%; Score 182.8; DB 9; Length 1734;
 Best Local Similarity 49.2%; Pred. No. 1,3e-43;
 Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

18 TCTGCCCAAGCATTTGATGAGGCGCTTCCGCAAGCGCCCTACCATGAGAGCGCCGT 77
 Db TTTTCTGATGATTTCACTTTGGGACAGCTGCTCGCGCTTTCAGTACGAAGGTGCAAC 149
 QY 78 CAAAGAAGTGGCGCGCGCCCTTCATCTGGACAGTACTGCACTGCAAGCCATGCGC 137
 Db 150 AAGTGAAGGTGGCAAGTCTCAACTATATGAGTACCTTCAAGCTCAGTACGAGAAG 209
 QY 138 CACCAAGCGCGCCCAAGCGCATGCTGCGATCACTACACGCGCTACGATGAGACTT 197
 Db 210 GACCAAAATGATATGATGAGATGATGATGATTTTATATATGATGATGATGATGAT 269
 QY 198 TGATCTTTGACCAAGTACGCGCAAGGCTACCGCTTCTCTGCTGCTGCTGCTGCT 257
 Db 270 AAATTTGATGAGAGCTAAACATGACGCTTCCGATTTTCAATCTCGTGTCAAGATT 329
 QY 258 CATTCCCTTGGCGGAGCGCTGATCCCTGCAACGAGAGGAAATGATTTTACAGCA 317
 Db 330 AATACCAGTGAAGAGCTAAAGATGAGTAAACAAAGAGGTGTACAAATTTCAAGGA 389
 QY 318 ACTGATTTGACGCGCTGTTGAGCGGCGGTATCAGCGCTTGGGTGACTTGTACCACTGGA 377
 Db 390 TCTCATATGAGCAACTTTCTTCTATATGATACATACAACTTTCATGACGCTTATCTTTGGA 449
 QY 378 TCTGCTTCAAGCGCTTCAAGCATGCTATGAGGCTGCTGCTGCTGCTGCTGCTGCT 437
 Db 450 CCAACCAATCTTTGAGAGCAAGATATGCTGCTTCTTCAAGCTTAAATTCGATGA--- 506
 QY 438 GGACTTTGAGCGGTATGCGAGGTTGCTTTGAAAGTTTGGGAGCCGATCCAGAACTG 497
 Db 507 AGATTTTCAGATTTTTCAGAAATTTGTTTGAAGATTTTGAAGATTTAAAGATGTG 586
 QY 498 GATCACCATCAAGCAAGCCCTGATTCAGGCACTATGATGATGATGATGATGATGATGAT 557
 Db 567 GACCAACATTAAGCAAGCTTATATATGATGATGATGATGATGATGATGATGATGATGAT 626
 QY 558 CCGCGGAGAGAGC-----AGCATTTAACAAGCACTCCACGAGGCAACCTGCACTGA 611
 Db 627 GCGTGAAGATGCTCAAAATGGGTAAAGAAAGTGTCAAGGCTGGAATTCGATGACGA 686
 QY 612 GCGGTGAGCTGCTGGAAGGCGCAAGATGATGAGGCACTGCGCGCGCTGCGCTTACAG 671
 Db 687 GCGTTACATTTGTTCAATCACTCACTCTTCTTCCCATGCGCTGCAAGTGAAGAAATTCG 746
 QY 672 CAGGACATTTGCGCCCTCGCAAAAGGCGCAAGTCCGATCTCGCTCAAGCGCACTACTA 731
 Db 747 AAAATGTGAAAAAACTTTCGATGATGAGGCAAAATGGGATGATGATGATGATGATGATGAT 806
 QY 732 TGAGCCCTGAGAGCAAGATGAGCTTCGGGCAAGAGAGCTGCTGAGCGAGTGAATTT 791
 Db 807 CGAGCTTATATCATTTCCGATTCACATGACATTAAGAGAGAGCTGGAAGAGAGCTTTCGCTT 866
 QY 792 TCACTTTGCTGTTGCTGCAATCTTCTTGAAGAGAGATCTATCAAGAGATGAA 851
 Db 867 TGAATTTGAGTGCATCTGATCAGTCAATTC--ACGAGATTTATCAAGAGATTTGAAA 923
 QY 852 GAAGGAGCTGGGAGAGAGCTTCAGGCTTCAGCTCCCGGAGAGCTTTGCCATTCCTAATGC 911
 Db 924 AAGTGAAGGTGGCAAGTCTTCACTTATTTACCTTTGAAACATTAAGATGTTACAAA 983
 QY 912 CGAGAGACCGACTTCTACGAGCATGATTTACTACATATCCAGTTTCCGCGCCCACTAGA 971
 Db 984 TTCA---TCAGATTTGCTGGAATTAATTAACAGGCTGCTGCTGCTGCTGCTGCTGCT 1040
 QY 972 CCGTCCCTGCGCGAGAGAGCATATCTGCGCGCATCATGA-----GACCAAGGA 1022
 Db 1041 TCACATAGACCTGAAAAAACTCGTTTCAAAATGACCACTGATGTGAATGAAACTGAC 1100

QY 1023 GAATTAAGACGAGCGCCCGTGTGCGAGAGAGAGCGGCTTCGCTGCTGCGCTCCTGCC 1082
 Db 1101 TAATCAACATGATGATCATATGATGATGATGATGATGATGATGATGATGATGATGAT 1160
 QY 1083 GGAATGTTTCCGAGAGCATCTCCCGCGGTGATGCGCTGTAC--GGCAAGCCATCTA 1139
 Db 1161 GGAAGGCTTACGAAAGTCTTAACTATATTAAGAGATGATGATGATGATGATGATGATGAT 1220
 QY 1140 CATCACGGAAGAGATGAGTCCCGCGCTGAGAGAGAGAAACATGACGAGGAGAGGCGCT 1199
 Db 1221 CATCAAGAAATGAAATTAATGATACACGAGCGGTACAAACAAAGAGAAATTTGT 1280
 QY 1200 CAAGACCCCTTCGCGATCGGATCTTGTGATCTGCACTTGAATTTTCAAGGCTAT 1259
 Db 1281 GAGGACACATTTAGATTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1340
 QY 1260 TACCAAGAGAGCGCTGCTGATCAAGGAGTATCTTGGTGGCGCTGCTGATTAATTGGA 1319
 Db 1341 AGTGAAGATGAGGTGATGAGTAAAGAGATTAATGATGATGATGATGATGATGATGATGAT 1400
 QY 1320 ATGTCAGATGCTAGGACCCGATTTGG 1349
 Db 1401 ATGGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1430

RESULT 15

US-09-938-842A-1282
 ; Sequence 1282, Application US/09938842A
 ; Publication No. US2004009476A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Krepe, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; FILE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SCRIPI300-3
 ; CURRENT APPLICATION NUMBER: US/09/938,842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264,647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300,111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 1282
 ; LENGTH: 1734
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-938-842A-1282

Query Match 12.6%; Score 182.8; DB 11; Length 1734;
 Best Local Similarity 49.2%; Pred. No. 1,3e-43;
 Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

18 TCTGCCCAAGCATTTGATGAGGCGCTTCCGCAAGCGCCCTACCAATGAGAGCGCCCGT 77
 Db TTTTCTGATGATTTCACTTTGGGACAGCTGCTCGCGCTTTCAGTACGAAGGTGCAAC 149
 QY 78 CAAAGAAGTGGCGCGCGCCCTTCATCTGGAGACATGATGCACTGCACTGAGAGCTGCGC 137
 Db 150 AAGTGAAGGTGGCAAGTCTTCACTATATGAGATGATGATGATGATGATGATGATGATGAT 209
 QY 138 CACCAAGCGCGCCCAAGCGGATGCTTGGCATCTACCAAGCTGATGAGAGACTT 197
 Db 210 GACCAAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 269
 QY 198 TGATCTTTGACCAAGTACGCGCAAGAGGCTTACCGCTTCTCTTGTGCTGCTGCGGAT 257
 Db 270 AAATTTGATGAAGAGCTAAACATGAGCGCTTTCGATTTTCAATCTCGTGTGCAAGATT 329
 QY 258 CATTCCCTTGGCGGAGCGCTGATCCGTCACAGAGAGAGAAATGATTTTACAGCAA 317

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 23:31:59 ; Search time 5013.54 Seconds
(Without alignments)
10211.501 Million cell updates/sec

Title: US-10-026-140-3
Perfect score: 1455
Sequence: 1 atgccagagtcctagctct.....99gtaagtcgacataa 1455

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Pending Patents_NA_Main:*

- 1: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 7: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 8: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 9: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 10: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 11: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 12: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 13: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
- 14: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 15: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 16: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 17: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 18: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 19: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 20: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 21: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 22: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 23: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 24: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 25: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 26: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 27: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 28: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 29: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 30: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 31: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 32: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 33: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 34: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 35: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 36: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 37: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 38: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 39: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 40: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 41: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 42: /cgn2_6/ptodata/2/pna/US09_COMB.seq:*
- 43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*

- 44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
- 45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
- 46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
- 47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
- 48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
- 49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
- 50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
- 51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
- 52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
- 53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
- 54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
- 55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
- 56: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
- 57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
- 58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
- 59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
- 60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
- 61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
- 62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
- 63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
- 64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
- 65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
- 66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
- 67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
- 68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
- 69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
- 70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
- 71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
- 72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
- 73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
- 74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
- 75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
- 76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
- 77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
- 78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
- 79: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
- 80: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
- 81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
- 82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
- 83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
- 84: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
- 85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
- 86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
- 87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
- 88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
- 89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
- 90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
- 91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
- 92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
- 93: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
- 94: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
- 95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
- 96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
- 97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
- 98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
- 99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
- 100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
- 101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
- 102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
- 103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
- 104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
- 105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
- 106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
- 107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
- 108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
- 109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
- 110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	1454.2	99.9	1455	1	PCT-US02-34674-3
2	1454.2	99.9	1455	1	PCT-US02-34764-3
3	1454.2	99.9	1455	43	US-10-026-140-3
4	1454.2	99.9	1455	1	PCT-US02-34674-1
5	1454.2	99.9	1455	1	PCT-US02-34764-1
6	1454.2	99.9	1455	43	US-10-026-140-1
7	598	41.1	735	22	US-09-533-559-7458
8	598	41.1	735	53	US-10-653-047-7458
9	345.4	23.7	1572	20	US-09-417-507-9340
10	345.4	23.7	1572	20	US-09-417-507-9340
11	318.6	21.9	2918	20	US-09-404-520-15403
12	317.8	21.8	1470	50	US-10-369-493-37092
13	317.8	21.8	1470	93	US-09-360-039-37092
14	307.6	21.1	2914	69	US-09-138-103-7477
15	294.6	20.2	1434	20	US-09-404-520-15403
16	276.6	19.0	2037	51	US-10-437-963-97407
17	266.6	18.3	1933	22	US-09-533-559-7405
18	266.6	18.3	1933	53	US-10-653-047-7405
19	249	17.1	1498	20	US-09-417-507-14756
20	246	16.9	1404	36	US-09-887-272A-4204
21	233.4	16.0	1688	47	US-10-219-999-10066
22	233.4	16.0	1688	51	US-10-425-114A-4348
23	233.4	16.0	1688	51	US-10-425-114A-4348
24	231.8	15.9	1711	30	US-09-708-427-79294
25	230.4	15.8	1794	47	US-10-219-999-14009
26	230.4	15.8	1794	51	US-10-425-114-19747
27	230.2	15.8	1794	51	US-10-425-114A-19747
28	227.4	15.6	1781	47	US-10-219-999-29391
29	227.4	15.6	1781	51	US-10-425-114-30939
30	227.4	15.6	1781	51	US-10-425-114A-30939
31	227.4	15.6	1781	51	US-10-425-114A-30939
32	227.4	15.6	1931	47	US-10-219-999-29394
33	227.4	15.6	1931	51	US-10-425-114-30835
34	227.4	15.6	1931	51	US-10-425-114A-30835
35	227.4	15.6	1932	47	US-10-219-999-16845
36	227.4	15.6	1932	51	US-10-425-114-33175
37	227.4	15.6	1932	51	US-10-425-114A-33175
38	227.4	15.6	1932	27	US-09-654-617-262950
39	227.4	15.6	1932	29	US-09-684-016-262950
40	219.2	15.1	1737	47	US-10-219-999-15047
41	219.2	15.1	1737	51	US-10-425-114-25892
42	219.2	15.1	1737	51	US-10-425-114A-25892
43	214.6	14.7	1503	36	US-09-887-272A-2847
44	214.6	14.7	1814	51	US-10-437-963-92182
45	206.8	14.2	1843	54	US-10-739-930-4492

ALIGNMENTS

RESULT 1
PCT-US02-34674-3
Sequence 3, Application PC/TUS0234674
GENERAL INFORMATION:
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Goedegebuur, Frits
APPLICANT: Ward, Michael
APPLICANT: Yao, Jian
TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
TITLE OF INVENTION: Encoding the Same
FILE REFERENCE: GCG97
CURRENT APPLICATION NUMBER: PCT/US02/34674
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1455
TYPE: DNA
ORGANISM: Trichoderma reesei

PCT-US02-34674-3					Query Match	99.9%	Score 1454.2	DB 1	Length 1455	
					Best Local Similarity 100.0%; Pred. No. 0;					
					Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	ATGCCGAGTGGCTGAGCTTGGCCCAACGATTTGATGAGGCTTGGCAAGCGCGCTTAC	60							
DB	1	ATGCCGAGTGGCTGAGCTTGGCCCAACGATTTGATGAGGCTTGGCAAGCGCGCTTAC	60							
QY	61	CAGATCAAGCCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	120							
DB	61	CAGATCAAGCCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	120							
QY	121	CACCTGAGCATGCGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	180							
DB	121	CACCTGAGCATGCGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	180							
QY	181	CGTACGATGAGGCTTGTATCTCTTGAACGATGACGCGCGCGCGCGCGCGCGCGCG	240							
DB	181	CGTACGATGAGGCTTGTATCTCTTGAACGATGACGCGCGCGCGCGCGCGCGCGCG	240							
QY	241	TTGCGGTGCG	300							
DB	241	TTGCGGTGCG	300							
QY	301	ATTGAGTTTACAGCAACTGATTTGACGCGCTGTTGAGCGCGGATACGCGCTTGGTG	360							
DB	301	ATTGAGTTTACAGCAACTGATTTGACGCGCTGTTGAGCGCGGATACGCGCTTGGTG	360							
QY	361	ACTTTGATCACTGGGATCTGCTCAGCGCTTCAAGATGCTATGAGGCTGGCTTAC	420							
DB	361	ACTTTGATCACTGGGATCTGCTCAGCGCTTCAAGATGCTATGAGGCTGGCTTAC	420							
QY	421	GTGGAAGGTCG	480							
DB	421	GTGGAAGGTCG	480							
QY	481	GACCGATTCAGCACTGATTCACATCAAGCCCTGATTCAGGCGCATATGATAT	540							
DB	481	GACCGATTCAGCACTGATTCACATCAAGCCCTGATTCAGGCGCATATGATAT	540							
QY	541	GCCACCGGAGCAAGCG	600							
DB	541	GCCACCGGAGCAAGCG	600							
QY	601	ACTGCCACTGAGCGGTGCTGCTGGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCG	660							
DB	601	ACTGCCACTGAGCGGTGCTGCTGGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCG	660							
QY	661	GCGGTCTACAGAGGAGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	720							
DB	661	GCGGTCTACAGAGGAGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	720							
QY	721	GCGCATCTATGAGCGCTGGGACAGCAATGAGCTCGGACAGAGGCGCTGAGCGA	780							
DB	721	GCGCATCTATGAGCGCTGGGACAGCAATGAGCTCGGACAGAGGCGCTGAGCGA	780							
QY	781	CGGATGGAATTTGACATTTGCTGTTGCAATCCCATCTTTGGAAGAGGACTATCCA	840							
DB	781	CGGATGGAATTTGACATTTGCTGTTGCAATCCCATCTTTGGAAGAGGACTATCCA	840							
QY	841	GAGAGCATGAAGAGGAGCGGAGGAGGCTTCAAGCGCTTCAAGCGGAGCTTGGCC	900							
DB	841	GAGAGCATGAAGAGGAGCGGAGGAGGCTTCAAGCGCTTCAAGCGGAGCTTGGCC	900							
QY	901	ATTCCTCATGCGGAGAGCGGAGGAGGCTTCAAGCGGAGGAGGCTTGGCC	960							
DB	901	ATTCCTCATGCGGAGAGCGGAGGAGGCTTCAAGCGGAGGAGGCTTGGCC	960							
QY	961	GCGCATCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1020							
DB	961	GCGCATCTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1020							

QY	1021	GAGAAATAAGACGGCAGCCCGCTTGAGGAGAGACGGCCCTCGCTCGCTCGCTCGC	1080
Db	1021	GAGAAATAAGACGGCAGCCCGCTTGAGGAGAGACGGCCCTCGCTCGCTCGCTCGC	1080
QY	1081	CCGACATGTGTTCCGAAGCATCTGCCCGGGGTGTACGGCTCTGTACCGGACAGCCCATCTAC	1140
Db	1081	CCGACATGTGTTCCGAAGCATCTGCCCGGGGTGTACGGCTCTGTACCGGACAGCCCATCTAC	1140
QY	1141	ATCACCGAGAACGGATGCCCCGTGCTCTGGAAGAGAAATGACGTGTGAGAGAGCCGTC	1200
Db	1141	ATCACCGAGAACGGATGCCCCGTGCTCTGGAAGAGAAATGACGTGTGAGAGAGCCGTC	1200
QY	1201	AACGACCCCTTCCGCATCCGCTACTTGTGACTCGCACTTGAGCTGATTTCCAAAGCCATT	1260
Db	1201	AACGACCCCTTCCGCATCCGCTACTTGTGACTCGCACTTGAGCTGATTTCCAAAGCCATT	1260
QY	1261	ACCCAGAGACGGCGTGTGCTGTCAGAGGGATCTTTGGTGTGGCGCTTGCTGATTACTTGGA	1320
Db	1261	ACCCAGAGACGGCGTGTGCTGTCAGAGGGATCTTTGGTGTGGCGCTTGCTGATTACTTGGA	1320
QY	1321	TGCTCAGATGCTTACGGACCCAGATTGGGCGCTACGTTCAACAGCTTACCAACCCTCAAG	1380
Db	1321	TGCTCAGATGCTTACGGACCCAGATTGGGCGCTACGTTCAACAGCTTACCAACCCTCAAG	1380
QY	1381	CGCAGCCGCAAGAGATCTGCCCTGCTCTCAAGACATGTTTGCGGCCCGCAGAGGCTT	1440
Db	1381	CGCAGCCGCAAGAGATCTGCCCTGCTCTCAAGACATGTTTGCGGCCCGCAGAGGCTT	1440
QY	1441	AAAGTGGCGGCATTA	1455
Db	1441	AAAGTGGCGGCATTA	1455

```

RESULT 2
PCT-US02-34764-3
Sequence 3, Application PC/TUS0234764
GENERAL INFORMATION:
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Goedegebuurt, Frits
APPLICANT: Ward, Michael
APPLICANT: Yao, Jian
TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
FILE REFERENCE: GC697
CURRENT APPLICATION NUMBER: PCT/US02/34764
CURRENT FILING DATE: 2003-10-30
NUMBER OF SEQ. ID NOS: 3
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ. ID NO 3
LENGTH: 1455
TYPE: DNA
ORGANISM: Trichoderma reesei
PCT-US02-34764-3

```

Query Match	99.9%	Score 1454.2	DB 1	Length 1455
Local Similarity	100.0%	Pred. NO. 0		
Matches 1455	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0

QY	1	ATGCCCGAGTGGCTAGCTCTGCCCCAAGAATTGAAATGGGGCTTCGCAACGGCGCCCTAC	60
Db	1	ATGCCCGAGTGGCTAGCTCTGCCCCAAGCACTTTGATGGGGCTTCGCAACGGCGCCCTAC	60
QY	61	CAGATCGAAGGCGCCGCTCAAGAAGAGTGGCGCGGCCCGCTCATCTGGAGACAGTATGCG	120
Db	61	CAGATCGAAGGCGCCGCTCAAGAAGAGTGGCGCGGCCCGCTCATCTGGAGACAGTATGCG	120
QY	121	CACCTGGAGCGATGGCGCGCACCAACGGCGCCAAACGGCGCATGTGGCTTGGGATACACTAC	180
Db	121	CACCTGGAGCGATGGCGCGCACCAACGGCGCCAAACGGCGCATGTGGCTTGGGATACACTAC	180
QY	181	CGCTACGATGAGGACTTTGATCTCTTGACCAAGTACGGCGCAAGAGGCTTACCGCTTCTCC	240

Db	181	CGCTACGATGAGGACTTTGATCTCTTGAACAAGTACGGCCGAAGGGCTACCGCTTCTCC	240
OY	241	TTGTCGTGTCGCGGATCATATCCCTCGCGCGGACTGATCCCTGTCACGAGAGGGA	300
Db	241	TTGTCGTGTCGCGGATCATATCCCTCGCGCGGACTGATCCCTGTCACGAGAGGGA	300
OY	301	ATTGAGTTTATACACAACTGATTGAGCGCTGTGTAAGCGGGGTATCACGCGCTTGGGTG	360
Db	301	ATTGAGTTTATACACAACTGATTGAGCGCTGTGTAAGCGGGGTATCACGCGCTTGGGTG	360
OY	361	ACTTTGTATACACTGGGATCTGCTCAGAGCGCTTACCATCGCTATGAGAGGTGACTCAAC	420
Db	361	ACTTTGTATACACTGGGATCTGCTCAGAGCGCTTACCATCGCTATGAGAGGTGACTCAAC	420
OY	421	GTGGAAGAGGTCCAGCTGGAATTGTAGCCGATATGCGAGGTGTGCTTTGAAAGTTTTGGG	480
Db	421	GTGGAAGAGGTCCAGCTGGAATTGTAGCCGATATGCGAGGTGTGCTTTGAAAGTTTTGGG	480
OY	481	GACCGAGTCCGAACTGGATACCATCAACGATCCCTGGAATTACAGGCTATATATGATAT	540
Db	481	GACCGAGTCCGAACTGGATACCATCAACGATCCCTGGAATTACAGGCTATATATGATAT	540
OY	541	GCCACCGGACGAAACGCCCCGGGAGAGACATTAACAGCATCTCACCGAGGGCAAC	600
Db	541	GCCACCGGACGAAACGCCCCGGGAGAGACATTAACAGCATCTCACCGAGGGCAAC	600
OY	601	ACTGCCACTGAGCCGTGGCTGCTGTGAAGAGGCCAGATCATGAGCCAATGCCCGCGCTG	660
Db	601	ACTGCCACTGAGCCGTGGCTGCTGTGAAGAGGCCAGATCATGAGCCAATGCCCGCGCTG	660
OY	661	GCCGTCTACAGCAGAGGACTTTGCGCCCTCGGAAAAGGGCCAGATCGGCTATTCCTCAAC	720
Db	661	GCCGTCTACAGCAGAGGACTTTGCGCCCTCGGAAAAGGGCCAGATCGGCTATTCCTCAAC	720
OY	721	GGCGACTACTATGAGCCCTGGGACAAGCAATGAGCTCGGGACAAGAGAGGCTGTGAGCGA	780
Db	721	GGCGACTACTATGAGCCCTGGGACAAGCAATGAGCTGTGGACAAGAGAGGCTGTGAGCGA	780
OY	781	CGGATGGAATTTCAATTTGCTGTTTGCCAMTCCATCTTCTTGAAGAGAGATATATCA	840
Db	781	CGGATGGAATTTCAATTTGCTGTTTGCCAMTCCATCTTCTTGAAGAGAGATATATCA	840
OY	841	GAGAGCATGAAGAGCAGCTGCGGAGAGGCTTCCAGCCTCACTCCCGCGGACTTTGGC	900
Db	841	GAGAGCATGAAGAGCAGCTGCGGAGAGGCTTCCAGCCTCACTCCCGCGGACTTTGGC	900
OY	901	ATCTCTCAATGCGGAGAGACCGACTTTCACGCGATGAAATTAATACATATCCCAATGGCG	960
Db	901	ATCTCTCAATGCGGAGAGACCGACTTTCACGCGATGAAATTAATACATATCCCAATGGCG	960
OY	961	CGCCACTGACGAGTCCCGTCCCGGAGACGGAATCTTCCGCGGCAATCCAGACACACG	1020
Db	961	CGCCACTGACGAGTCCCGTCCCGGAGACGGAATCTTCCGCGGCAATCCAGACACACG	1020
OY	1021	GAGAAATAAGACGCGCAGCCCGTTGGGCGAGGAGCGGCTTCGCTGCGTGGCTCTGC	1080
Db	1021	GAGAAATAAGACGCGCAGCCCGTTGGGCGAGGAGCGGCTTCGCTGCGTGGCTCTGC	1080
OY	1081	CCGGACATGTTCCGGAAAGCATCTGCGCCCGGATGTAAGGCTGTACGGCAACCCATATCAC	1140
Db	1081	CCGGACATGTTCCGGAAAGCATCTGCGCCCGGATGTAAGGCTGTACGGCAACCCATATCAC	1140
OY	1141	ATCACCGAGAACGATGCGCGTGCCTGTGAGAGAGAGAAATGATGCTGCGAGAGAGCGCTC	1200
Db	1141	ATCACCGAGAACGATGCGCGTGCCTGTGAGAGAGAGAAATGATGCTGCGAGAGAGCGCTC	1200
OY	1201	AACGACCCCTTCCGCAATCCGTAATTTGACTCGCACTTGGACTCGATTTTCCAAAGCCATT	1266
Db	1201	AACGACCCCTTCCGCAATCCGTAATTTGACTCGCACTTGGACTCGATTTTCCAAAGCCATT	1266
OY	1261	ACCGACGAGCGGCGTCTGCTCAAGGGATATTTGGCGGGCGTTGCTCGATATCTTGAA	1320
Db	1261	ACCGACGAGCGGCGTCTGCTCAAGGGATATTTGGCGGGCGTTGCTCGATATCTTGAA	1320

```
QY 1321 TGTTCAGATGGCTACGAGCCGAGATTGGCGGTCACGTTACAGACTACACCACTCTAG 1380
DB 1321 TGTTCAGATGGCTACGAGCCGAGATTGGCGGTCACGTTACAGACTACACCACTCTAG 1380
QY 1381 CGCAGCGCCCAAGACTGTGCTTGTCTCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440
DB 1381 CGCAGCGCCCAAGACTGTGCTTGTCTCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440
QY 1441 AAAGTGGCGGCATTA 1455
DB 1441 AAAGTGGCGGCATTA 1455

RESULT 3
US-10-026-140-3
; Sequence 3, Application US/10026140
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: US/10/026,140
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-140-3

Query Match 99.9%; Score 1454.2; DB 43; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCGAGTGTGCTAGCTTGTGCTCCCAAGACTTGTGAGTGGGCTTCGCAAGCGGCGCTTAC 60
DB 1 ATGCCGAGTGTGCTAGCTTGTGCTCCCAAGACTTGTGAGTGGGCTTCGCAAGCGGCGCTTAC 60
QY 61 CAGATCGAAGGCGCGTCTCAAGAAAGTGGCGCGCGCTTCATCTGGGACACGTACTGC 120
DB 61 CAGATCGAAGGCGCGTCTCAAGAAAGTGGCGCGCGCTTCATCTGGGACACGTACTGC 120
QY 121 CACCTGAGGCGATGCGGCGCAACAGCGGCGCAAGTGGCTTGGCATCTACAC 180
DB 121 CACCTGAGGCGATGCGGCGCAACAGCGGCGCAAGTGGCTTGGCATCTACAC 180
QY 181 CGCTACGATGAGGACTTTGATCTCTTGACCAAGTACGCGCAAGGCGCTACCGCTTCTC 240
DB 181 CGCTACGATGAGGACTTTGATCTCTTGACCAAGTACGCGCAAGGCGCTACCGCTTCTC 240
QY 241 TTTGTCGTGTGTCGGGATCATTTCCCTTGGCGGAGGTGATCCCGTCAAGAGGGA 300
DB 241 TTTGTCGTGTGTCGGGATCATTTCCCTTGGCGGAGGTGATCCCGTCAAGAGGGA 300
QY 301 ATTGAGTTTACAGCAAACTGATGAGCGGCTGTGAGGGGGGTATCAAGCTTGGGTG 360
DB 301 ATTGAGTTTACAGCAAACTGATGAGCGGCTGTGAGGGGGGTATCAAGCTTGGGTG 360
QY 361 ACTTTGACACTGAGATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGGCTAAC 420
DB 361 ACTTTGACACTGAGATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGGCTAAC 420
QY 421 GTGGAGAGGTTCCAGCTGAGCTTTGAGCGGTATGCGAGTGTGCTTTGAACGTTTGGG 480
DB 421 GTGGAGAGGTTCCAGCTGAGCTTTGAGCGGTATGCGAGTGTGCTTTGAACGTTTGGG 480
QY 481 GACCGAGTCCAGAACTGATCACCATCAAGWCCGATGAGGCAATCTATGATAT 540
DB 481 GACCGAGTCCAGAACTGATCACCATCAAGWCCGATGAGGCAATCTATGATAT 540
```

```
DB 481 GACCGAGTCCAGAACTGATCACCATCAAGWCCGATGAGGCAATCTATGATAT 540
QY 541 GCCACCGGACGCAACGCTCCCGGAGAGGAGCATTTAACAGACTTCCACGAGGGCAAC 600
DB 541 GCCACCGGACGCAACGCTCCCGGAGAGGAGCATTTAACAGACTTCCACGAGGGCAAC 600
QY 601 ACTGCCACTGAGCGGTGCTGCTGGAAAGGCCCAATCATGAGCCATGCTCCGCGCTG 660
DB 601 ACTGCCACTGAGCGGTGCTGCTGGAAAGGCCCAATCATGAGCCATGCTCCGCGCTG 660
QY 661 GCGGTCTACAGCAGGACTTTCGCTCCCTCGCAAAAGGCGCAGATCCGATCTGCTAAC 720
DB 661 GCGGTCTACAGCAGGACTTTCGCTCCCTCGCAAAAGGCGCAGATCCGATCTGCTAAC 720
QY 721 GCGGACTACTATGAGCCTTGGGACAGCAATGAGCTTCGGGACAAAGAGCTGCTGAGCGA 780
DB 721 GCGGACTACTATGAGCCTTGGGACAGCAATGAGCTTCGGGACAAAGAGCTGCTGAGCGA 780
QY 781 CGGATGGAATTTCAATTTGGCTGTGTCGCAATCCCATCTTTGAAAGAGACTATCCA 840
DB 781 CGGATGGAATTTCAATTTGGCTGTGTCGCAATCCCATCTTTGAAAGAGACTATCCA 840
QY 841 GAGAGCATGAAGAGAGCTGGGCGAGAGGCTTTCAGCCTCACTCCCGGAGCTTTGCGC 900
DB 841 GAGAGCATGAAGAGAGCTGGGCGAGAGGCTTTCAGCCTCACTCCCGGAGCTTTGCGC 900
QY 901 ATCCTCAAGCGCGGAGAGACCGACTTCTACGGATGAATTAATCAATCCAGTTTCGG 960
DB 901 ATCCTCAAGCGCGGAGAGACCGACTTCTACGGATGAATTAATCAATCCAGTTTCGG 960
QY 961 CGCAGCTTGAAGCGGTCCTCCGTCGCGAGAGGACTATCTGCGGCGCATTCATGAGCAC 1020
DB 961 CGCAGCTTGAAGCGGTCCTCCGTCGCGAGAGGACTATCTGCGGCGCATTCATGAGCAC 1020
QY 1021 GAGATTAAGAGCGGACGCTTGGCGAGAGAGCGGCTTGTGCTGCTGCTTGC 1080
DB 1021 GAGATTAAGAGCGGACGCTTGGCGAGAGAGCGGCTTGTGCTGCTGCTTGC 1080
QY 1081 CGGACATGTTCCGAGAGCATCTGCGCGGAGTACGCGCTGTAACGCGCAAGCCATCTAC 1140
DB 1081 CGGACATGTTCCGAGAGCATCTGCGCGGAGTACGCGCTGTAACGCGCAAGCCATCTAC 1140
QY 1141 ATCAGCGAGAGCGAGTCCGCTGCTGAGAGAGGAGCATGATGATGAGAGGCGCTC 1200
DB 1141 ATCAGCGAGAGCGAGTCCGCTGCTGAGAGAGGAGCATGATGATGAGAGGCGCTC 1200
QY 1201 AAGAGCCCTTCCGATCCGATCTTGAATCTGCACTTGAATCTGATTTCCAAAGCAT 1260
DB 1201 AAGAGCCCTTCCGATCCGATCTTGAATCTGCACTTGAATCTGATTTCCAAAGCAT 1260
QY 1261 ACCCAGAGCGGCGTCTGCAAGAGGATCTTGTGCTGGGCTGTGCTGATTAATCTTGA 1320
DB 1261 ACCCAGAGCGGCGTCTGCAAGAGGATCTTGTGCTGGGCTGTGCTGATTAATCTTGA 1320
QY 1321 TGTTCAGATGGCTACGAGCCGAGATTGGCGGTCACGTTACAGACTACACCACTCTAG 1380
DB 1321 TGTTCAGATGGCTACGAGCCGAGATTGGCGGTCACGTTACAGACTACACCACTCTAG 1380
QY 1381 CGCAGCGCCCAAGACTGTGCTTGTCTCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440
DB 1381 CGCAGCGCCCAAGACTGTGCTTGTCTCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440
QY 1441 AAAGTGGCGGCATTA 1455
DB 1441 AAAGTGGCGGCATTA 1455

RESULT 4
PCT-US02-34674-1
; Sequence 1, Application PC/TUS0234674
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
```

APPLICANT: Ward, Michael
APPLICANT: Yao, Jian
TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
TITLE OF INVENTION: Encoding the Same
FILE REFERENCE: GC697
CURRENT APPLICATION NUMBER: PCT/US02/34674
CURRENT FILING DATE: 2003-03-14
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1991
TYPE: DNA
ORGANISM: Trichoderma reesei
PCT-US02-34674-1

Query Match 99.9%; Score 1454.2; DB 1; Length 1991;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCGAGTCGCTAGCTGCTGCCAAGCACTTGAATGGGGCTTCGCAACGGCCGCTTAC 60
DB 62 ATGCCGAGTCGCTAGCTGCTGCCAAGCACTTGAATGGGGCTTCGCAACGGCCGCTTAC 121
QY 61 CAGATCGAAGCGCCGCTCAAGAAAGTGCGCGCCGCTTCATCTGGACACGTAATGCG 120
DB 122 CAGATCGAAGCGCCGCTCAAGAAAGTGCGCGCCGCTTCATCTGGACACGTAATGCG 181
QY 121 CAGCTGAGGCGATTCGCGCAGCAACGGCGCCCAAGGCGATGGCTTCGCACTTACAC 180
DB 182 CAGCTGAGGCGATTCGCGCAGCAACGGCGCCCAAGGCGATGGCTTCGCACTTACAC 241
QY 181 CGCTACGATGAGGACTTGTGATCTCTTGAACCAAGTACGGCGCAAGGCGCTACCGCTTCTCC 240
DB 242 CGCTACGATGAGGACTTGTGATCTCTTGAACCAAGTACGGCGCAAGGCGCTACCGCTTCTCC 301
QY 241 TTGTCGTGTCGCGCGATCATTTCCCTCGCGGCGAGGCTGTGATCCGCTCAAGGAGAGGGA 300
DB 302 TTGTCGTGTCGCGCGATCATTTCCCTCGCGGCGAGGCTGTGATCCGCTCAAGGAGAGGGA 361
QY 301 ATTGAGCTTTTACAGCAAACTGATTTGAGCGCTTGTGAGGGGGATTTACAGCTTTGGGTG 360
DB 362 ATTGAGCTTTTACAGCAAACTGATTTGAGCGCTTGTGAGGGGGATTTACAGCTTTGGGTG 421
QY 361 ACTTTGACCACTGGGATCTGCTCAGGCGCTTACAGTGGCTTGTGAGGGCTGCTGAC 420
DB 422 ACTTTGACCACTGGGATCTGCTCAGGCGCTTACAGTGGCTTGTGAGGGCTGCTGAC 481
QY 421 GTGGAAGAGTCCAGCTGGAATTGAGCGGTATCGAGGTTGTGCTTTGAAAGTTTTGGG 480
DB 482 GTGGAAGAGTCCAGCTGGAATTGAGCGGTATCGAGGTTGTGCTTTGAAAGTTTTGGG 541
QY 481 GACCGAGTCCAGAACTGATATCAATCAAGCAAGCCCTGATTTACGGCCATTTATGATAT 540
DB 542 GACCGAGTCCAGAACTGATATCAATCAAGCAAGCCCTGATTTACGGCCATTTATGATAT 601
QY 541 GCCACGGGAGCAAGCGCCGGGAGAGAGCATTTAAAGCACTCCACCGAGGGGAAC 600
DB 602 GCCACGGGAGCAAGCGCCGGGAGAGAGCATTTAAAGCACTCCACCGAGGGGAAC 661
QY 601 ACTGCCACTGAGCCGTGCTGCTGTAAGGCGCAAGTCAATGAGCCATGCGCGCGCTG 660
DB 662 ACTGCCACTGAGCCGTGCTGCTGTAAGGCGCAAGTCAATGAGCCATGCGCGCGCTG 721
QY 661 GCCGCTTACAGCAGGAGACTTTCCGCTCCGCAAAAAGGCGCAAGTGGCGATTCGCTCAAC 720
DB 722 GCCGCTTACAGCAGGAGACTTTCCGCTCCGCAAAAAGGCGCAAGTGGCGATTCGCTCAAC 781
QY 721 GGGGACTACTATGAGCCCTGGGAGCAGCAATGAGCTGGGAGCAAGAGGCTGTGAGGGA 780
DB 782 GGGGACTACTATGAGCCCTGGGAGCAGCAATGAGCTGGGAGCAAGAGGCTGTGAGGGA 841
QY 781 CGGATGGAATTTCACTATGGCTGCTTGGCAATCCATCTTTTGAAGAGACTATATCA 840

DB 842 CGGATGGAATTTCACTATGGCTGCTTGGCAATCCATCTTTTGAAGAGACTATATCA 901
QY 841 GAGAGCATGAAGAGAGCTGGGAGAGGCTTCAAGCCCTCACTCCCGGAGCTTTGGC 900
DB 902 GAGAGCATGAAGAGAGCTGGGAGAGGCTTCAAGCCCTCACTCCCGGAGCTTTGGC 961
QY 901 ATCTCAATGCGCGAGAGACCGACTTCTACGAGATGAATTTACTACATCCAGTTGGCG 960
DB 962 ATCTCAATGCGCGAGAGACCGACTTCTACGAGATGAATTTACTACATCCAGTTGGCG 1021
QY 961 CGCCACTTACAGCGCTCCGCTCCCGAGAGCGGACTATTTGGCGCCATCCATGAGCACAG 1020
DB 1022 CGCCACTTACAGCGCTCCGCTCCCGAGAGCGGACTATTTGGCGCCATCCATGAGCACAG 1081
QY 1021 GAGATTAAGAGCGGAGCGCCGCTTGGGAGAGAGGCGGCTTCGCTGGCTGCTGCG 1080
DB 1082 GAGATTAAGAGCGGAGCGCCGCTTGGGAGAGAGGCGGCTTCGCTGGCTGCTGCG 1141
QY 1081 CCGGACATGTTCCGGAAGCATCTTCGCGCGGCTGTACGCGCTGTACGCAAGCCCATCTAC 1140
DB 1142 CCGGACATGTTCCGGAAGCATCTTCGCGCGGCTGTACGCGCTGTACGCAAGCCCATCTAC 1201
QY 1141 ATCAACGAGAAAGGATGCGGCTGCTGAGAGAGAGAACTGACGTGCGAGAGCGCTG 1200
DB 1202 ATCAACGAGAAAGGATGCGGCTGCTGAGAGAGAGAACTGACGTGCGAGAGCGCTG 1261
QY 1201 AAGGACCCCTTCGCGATCCGCTACTTGTGATCCGACTTGGACTTGATTTCAAGGCGCAT 1260
DB 1262 AAGGACCCCTTCGCGATCCGCTACTTGTGATCCGACTTGGACTTGATTTCAAGGCGCAT 1321
QY 1261 ACCGAGACCGGCTGCTGCTCAAGGGGTAATTTGCGTGGCGTGTGCTGATTAATTGGAA 1320
DB 1322 ACCGAGACCGGCTGCTGCTCAAGGGGTAATTTGCGTGGCGTGTGCTGATTAATTGGAA 1381
QY 1321 TGGTCAGATGCTTACCGAGCCCAATTTGGCGGTCACTTACAGACTTACACACCTTCAAG 1380
DB 1382 TGGTCAGATGCTTACCGAGCCCAATTTGGCGGTCACTTACAGACTTACACACCTTCAAG 1441
QY 1381 CGGACCGCCCAAGAAAGTGTGCGCTGCTCAAGAGCAATGTTGCGCGCGGAGAGGGTT 1440
DB 1442 CGGACCGCCCAAGAAAGTGTGCGCTGCTCAAGAGCAATGTTGCGCGCGGAGAGGGTT 1501
QY 1441 AAAGTGGCGGCAATTA 1455
DB 1502 AAAGTGGCGGCAATTA 1516

RESULT 5
PCT-US02-34764-1

Sequence 1, Application PC/TUS0234764
GENERAL INFORMATION:
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Goedegebuur, Frits
APPLICANT: Ward, Michael
APPLICANT: Yao, Jian
TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
TITLE OF INVENTION: Encoding the Same
FILE REFERENCE: GC697
CURRENT APPLICATION NUMBER: PCT/US02/34764
CURRENT FILING DATE: 2003-10-30
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1991
TYPE: DNA
ORGANISM: Trichoderma reesei
PCT-US02-34764-1

Query Match 99.9%; Score 1454.2; DB 1; Length 1991;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCCGAGTCGCTAGCTGCTGCCAAGCACTTGAATGGGGCTTCGCAACGGCCGCTTAC 60

|||||
Db 362 ATTGAGTTTAAAGCAAACTGATTTGAGCGCCTGTTGAGCGGGGTATACGCGCTTGGGTG 421
Qy 361 ACTTGTACACATGGGATCTGCCCTCAGCGCTTCAGATGCGTATGAGGGTGGCTAAC 420
Db 422 ACTTGTACACATGGGATCTGCCCTCAGCGCGCTTCAGATGCGTATGAGGGTGGCTAAC 481
Qy 421 GTGGAAGAGGTCCAGCTGGAATTTGAGCGGATGCGAGGTTGTGCTTGAACGTTTTGGG 480
Db 482 GTGGAAGAGGTCCAGCTGGAATTTGAGCGGATGCGAGGTTGTGCTTGAACGTTTTGGG 541
Qy 481 GACCGAGTCCAGAACTGGAATCACCATTACAGAACCTTGTGATTCAGGCCATCTATGATAT 540
Db 542 GACCGAGTCCAGAACTGGAATCACCATTACAGAACCTTGTGATTCAGGCCATCTATGATAT 601
Qy 541 GGCAACGGGCGCAACGGCGCGGAGAGAGCACTTAACAAGCACTCCACGAGGGCAAC 600
Db 602 GGCAACGGGCGCAACGGCGCGGAGAGAGCACTTAACAAGCACTCCACGAGGGCAAC 661
Qy 601 ACTGCCACTGAGCGGTGGCTCGCTGAAAGGCCAGATCATGAGCCATGCGCGCGCGCTG 660
Db 662 ACTGCCACTGAGCGGTGGCTCGCTGAAAGGCCAGATCATGAGCCATGCGCGCGCGCTG 721
Qy 661 GCCGTCTACAGCAGGGAATTTCCGCTTCCGAAAAGGCCAGATCGGCATCTGCTCAAC 720
Db 722 GCCGTCTACAGCAGGGAATTTCCGCTTCCGAAAAGGCCAGATCGGCATCTGCTCAAC 781
Qy 721 GGCGACTACTATAGCGCTGGGAGAGCAATGAGCGCTGGGAGAGAGGCTGCTGAGCGA 780
Db 782 GGCGACTACTATAGCGCTGGGAGAGCAATGAGCGCTGGGAGAGAGGCTGCTGAGCGA 841
Qy 781 CGGATGGAATTTTCATTTGGCTGGTTTGCAGATCCCATCTTCTTGAAGAAGATATCCA 840
Db 842 CGGATGGAATTTTCATTTGGCTGGTTTGCAGATCCCATCTTCTTGAAGAAGATATCCA 901
Qy 841 GAGAGCATGAAGAAGCAGCTGGGCGAGAGGCTTCAGCGCTCACTCCCGGAGATTTGCC 900
Db 902 GAGAGCATGAAGAAGCAGCTGGGCGAGAGGCTTCAGCGCTCACTCCCGGAGATTTGCC 961
Qy 901 ATCTCTCAATGCCGAGAGACCGCATTTTACAGGCAATTAATCAATCCCACTGTTGGC 960
Db 962 ATCTCTCAATGCCGAGAGACCGCATTTTACAGGCAATTAATCAATCCCACTGTTGGC 1021
Qy 961 CGGCACCTTACAGCGGTCCCGTCCCGAGAGCACTATCTCGCGCGCATCCATGAGCAG 1020
Db 1022 CGGCACCTTACAGCGGTCCCGTCCCGAGAGCACTATCTCGCGCGCATCCATGAGCAG 1081
Qy 1021 GAGAAATTAAGACGCGCAGCGCGCTTGGCGAGAGAGCGCGCTCGCTGGCTCTGC 1080
Db 1082 GAGAAATTAAGACGCGCAGCGCGCTTGGCGAGAGAGCGCGCTCGCTGGCTCTGC 1141
Qy 1081 CGCGACATGTTCCCGGAGAGCATCTCGCGCGGATGACCGGCTGTAACCGCAAGCCATCTAC 1140
Db 1142 CGCGACATGTTCCCGGAGAGCATCTCGCGCGGATGACCGGCTGTAACCGCAAGCCATCTAC 1201
Qy 1141 ATCAACGAGAAGGATGCGCGCTGAGAGAGAGCACTGACGTGCGAGAGAGCGCTC 1200
Db 1202 ATCAACGAGAAGGATGCGCGCTGAGAGAGAGCACTGACGTGCGAGAGAGCGCTC 1261
Qy 1201 AACGACCCCTTCCGATCCGTAATTTGACTCGCACTTGAATCTGATTTTCAAGGCCAT 1260
Db 1262 AACGACCCCTTCCGATCCGTAATTTGACTCGCACTTGAATCTGATTTTCAAGGCCAT 1321
Qy 1261 ACCGAGAGCGCGCTGCTGTCMAAGGGTACTTTGCGTGGCGCTTGTCTGATTAATTTGGA 1320
Db 1322 ACCGAGAGCGCGCTGCTGTCMAAGGGTACTTTGCGTGGCGCTTGTCTGATTAATTTGGA 1381
Qy 1321 TGGTCAGATGAGCTACGAGACCAATTCGCGGTCACTTCAAGACTACACCAAGCTTCAAG 1380
Db 1382 TGGTCAGATGAGCTACGAGACCAATTCGCGGTCACTTCAAGACTACACCAAGCTTCAAG 1441
Qy 1381 CGGACGCCCAAGAGTCTGCTGCTGCTCAAGGACATGTTTGGCGCGCGGAGAGGGTT 1440
|||||

Db 1442 CGGACGCCCAAGAGTCTGCGCTGCTGTCCTTCAGAGACATGTTTGGCGCGGAGAGGGTT 1501
Qy 1441 AAAGTGCGGCGCATTA 1455
Db 1502 AAAGTGCGGCGCATTA 1516
RESULT 7
US-09-533-559-7458
Sequence 7458, Application US/09533559
GENERAL INFORMATION:
APPLICANT: Randy M. Berka
APPLICANT: Michael W. Rey
APPLICANT: Jeffrey R. Shuster
APPLICANT: Sakari Kauppinen
APPLICANT: Ib Groth Clausen
APPLICANT: Peter Bjarke Olsen
TITLE OF INVENTION: Methods for Monitoring Multiple Gene
FILE REFERENCE: 5849.200-US
CURRENT APPLICATION NUMBER: US/09/533,559
EARLIER FILING DATE: 2000-03-22
EARLIER APPLICATION NUMBER: 09/273,623
NUMBER OF SEQ ID NOS: 7860
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7458
LENGTH: 735
TYPE: DNA
ORGANISM: Tricoderma reesei
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) - (735)
OTHER INFORMATION: n = A,T,C or G
US-09-533-559-7458
Query Match 41.1%; Score 598; DB 22; Length 735;
Best Local Similarity 97.0%; Pred. No. 1,4e-126;
Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;
Qy 258 CATTCCCTCGCGGAGCGCTGATCCGTCACAGAGAGGAATTTGATTTTACAGCAA 317
Db 1 CATTCCCTCGCGGAGCGCTGATCCGTCACAGAGAGGAATTTGATTTTACAGCAA 60
Qy 318 ACTGATTGAGCGCCTGTGAGGCGGGGTATCAAGCCTTGGGTGATCTTGTACCACTGGGA 377
Db 61 ACTGATTGAGCGCCTGTGAGGCGGGGTATCAAGCCTTGGGTGATCTTGTACCACTGGGA 120
Qy 378 TCTGCTCAGGCGCTTCACGATGCTATGAGGCTGGCTCAACGTGAAAGAGTCCAGCT 437
Db 121 TCTGCTCAGGCGCTTCACGATGCTATGAGGCTGGCTCAACGTGAAAGAGTCCAGCT 180
Qy 438 GAACTTTGAGCGGTATGCGAGGTTGTGCTTTGAACGTTTTTGGGACCGAGTCCAGAACTG 497
Db 181 GAACTTTGAGCGGTATGCGAGGTTGTGCTTTGAACGTTTTTGGGACCGAGTCCAGAACTG 240
Qy 498 GATCACCATTAAGAACCCCTGATTCAGGCACTATGATATGATATGCCAGCGGCAACGC 557
Db 241 GATCACCATTAAGAACCCCTGATTCAGGCACTATGATATGATATGCCAGCGGCAACGC 300
Qy 558 CCGGCGAGAGGAGATTAACAAGCACTCAACGAGGGCAACTGCGCACTGAGCGGTG 617
Db 301 CCGGCGAGAGGAGATTAACAAGCACTCAACGAGGGCAACTGCGCACTGAGCGGTG 360
Qy 618 GCTTGTGAAAGGCCCAAGATCATGAGCCATGCGCGCGCTGTGCTTACAGCAGGGA 677
Db 361 GCTTGTGAAAGGCCCAAGATCATGAGCCATGCGCGCGCTGTGCTTACAGCAGGGA 420
Qy 678 CTTTGGCCCTTGCAGAAAGGCGAGATCGGCATCTGCTCAACGAGGCACTATGAGCC 737
Db 421 CTTTGGCCCTTGCAGAAAGGCGAGATCGGCATCTGCTCAACGAGGCACTATGAGCC 480
Qy 738 CTGGGACAGCAATGAGCTTCGGAGCAAGAGAGCTGTGAGCGAGCATGATTAATTTCAAT 797
|||||


```
Db 481 CTTGGAGAGCATATAGCTCCGGACAAAGAGGCTGTGAGGAGAGATGAAATTTTCAT 540
Qy 798 TGGCTGTTGGCAATCCCATCTTCTTGAAGAAGACTA-TCCAGAGAGCATGAAGAAGC 856
Db 541 TGGCTGTTGGCAATCCCATCTTCTTGAAGAAGACTA-TCCAGAGAGCATGAAGAAGC 600
Qy 857 AGCT--GGGCGAGAGGCTTCAGGCTTCAC-TCCCGGAGCTTT--GCCATCTCAATGC 911
Db 601 AACTTGGGGCGAGAGGCTTTCAACCTTTACTTCCCGGAGACTTTGCGCATCTTMAATGC 660
Qy 912 CGGAGAGAC 921
Db 661 CGGAGAGAAC 670

RESULT 8
US-10-653-047-7458
; Sequence 7458, Application US/10653047
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjørke Olsen
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/10/653,047
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US/09/533,559
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/273,623
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7458
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Tricoderma reesei
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(735)
; OTHER INFORMATION: n = A,T,C or G
US-10-653-047-7458

Query Match 41.1%; Score 598; DB 53; Length 735;
Best Local Similarity 97.0%; Pred. No. 1.4e-126;
Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;

Qy 258 CATTCCCTCGGGGAGAGGCTGATCCCGTCAACGAGAGGAGGAAATTTAAGCA 317
Db 1 CATTCCCTCGGGGAGAGGCTGATCCCGTCAACGAGAGGAGGAAATTTAAGCA 60
Qy 318 ACTGATGAGCGCCCTGTGAGGCGGGGTATCAAGCCTTGGGTGACTTTTACCACTGGA 377
Db 61 ACTGATGAGCGCCCTGTGAGGCGGGGTATCAAGCCTTGGGTGACTTTTACCACTGGA 120
Qy 378 TCTGCTCAGGCGCTTACAGATGCTATGAGAGGCTGCTCAACGTGGAAGAGTCCAGCT 437
Db 121 TCTGCTCAGGCGCTTACAGATGCTATGAGAGGCTGCTCAACGTGGAAGAGTCCAGCT 180
Qy 438 GGACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGAGCGAGTCCAGAACTG 497
Db 181 GGACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGAGCGAGTCCAGAACTG 240
Qy 498 GATCAACATCAAGAMCCTGTGATTCAAGGCACTATGATATGCCACCGGAGCAACGC 557
Db 241 GATCAACATCAAGAACCTGTGATTCAAGGCACTATGATATGCCACCGGAGCAACGC 300
Qy 558 CCGGAGAGAGAGCATTTAACAAGACTCAACGAGGAGCAACACTGCCATGAGCGGTG 617
```

```
Db 301 CCGGAGAGAGAGCATTTAAACAAGACTCCACGAGGAGCAACACTGCCACTGAGCCGTG 360
Qy 618 GCTCGCTGAAAAGGCCCAATATGATGCAATGCCCGGCGCGGTCTACAGAGGGA 677
Db 361 GCTCGCTGAAAAGGCCCAATATGATGCAATGCCCGGCGCGGTCTACAGAGGGA 420
Qy 678 CTTTGGCCCTCGCAAAAAGGCGAGATCGGACTTTCGCTCAACGGGCACTACTATGAGCC 737
Db 421 CTTTGGCCCTCGCAAAAAGGCGAGATCGGACTTTCGCTCAACGGGCACTACTATGAGCC 480
Qy 738 CTGGAAGCAATAGAGCTTCGGGACAAAGAGGCTGTGAGCGAGCGAATGGAATTTCAAT 797
Db 481 CTGGAAGCAATAGAGCTTCGGGACAAAGAGGCTGTGAGCGAGCGAATGGAATTTCAAT 540
Qy 798 TGGCTGTTGGCAATCCCATCTTCTTGAAGAAGACTA-TCCAGAGAGCATGAAGAAGC 856
Db 541 TGGCTGTTGGCAATCCCATCTTCTTGAAGAAGACTA-TCCAGAGAGCATGAAGAAGC 600
Qy 857 AGCT--GGGCGAGAGGCTTCAGGCTTCAC-TCCCGGAGCTTT--GCCATCTCAATGC 911
Db 601 AACTTGGGGCGAGAGGCTTTCAACCTTTACTTCCCGGAGACTTTGCGCATCTTMAATGC 660
Qy 912 CGGAGAGAC 921
Db 661 CGGAGAGAAC 670

RESULT 9
US-09-417-507-9340
; Sequence 9340, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; TITLE OF INVENTION: FUMIGATUS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PAT99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 9340
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: A.fumigatus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1562)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-417-507-9340

Query Match 23.7%; Score 345.4; DB 20; Length 1572;
Best Local Similarity 55.7%; Pred. No. 1.4e-68;
Matches 721; Conservative 2; Mismatches 563; Indels 9; Gaps 3;

Qy 27 CGACTTTGATGAGGCTTTCGCAACGCGCGCTTACAGATCGAAGGCGCCGTAAGAAGG 86
Db 66 CGACTTTTTCATGATTAAGCAACGCGCGCGCGCAAGTGAAGAGCTTGAATAAGA 125
Qy 87 TGGCGCGCGCGCTTCATCTGGGACAGTACTGCCACTGAGGCAATCGGCAACAAG 146
Db 126 CGTAAAGGCCAATCAATCTGGGATCTTTTGGCACAACCTGGGAAGGTGAAGAGCG 185
Qy 147 CGGCAAGGCGATGAGGCTTCGATCACTACACCGCTACGATGAGACTTTGATCTCT 206
Db 186 AAGCATGAGGAGCAAGCGGGTACGCTGTATGACTCTGTACAAAGAAATGTGGGCTCAT 245
Qy 207 GACCAAGTACGCGCAAGGCGCTTACCGCTTCTGTGTGTGCGGATCATTTCCCT 266
Db 246 GAAGTGTATGAGGAGTCAATGATACCGCTTCTCTGTGTGTGCTGCAATCATTTCTCT 305
Qy 267 CGGCGGAGGCTGATCCGCTCAACGAGAGGAAATGAGTTTACAGCAACATGATGA 326
Db 306 TGGAGGGTGCAGCAGCTTCGCAAGAAAGGAGATTAATTAATTTCAATTTGTTGA 365
Qy 327 CGGCTGTTGAGCGGGGTATCAAGCTTGGGTGACTTTGTATCAACTGGGATCTGCTCA 386
```

```

Db      366 CGAGCTTCTCCGTAATGATCATACACCGTTTGTCTACTCTCTCCACATGGGACACACGCA 425
Qy      387 GGGGCTTACAGATTCGCTATGAGAGCTGGCTCAACGTGGAAGAGTCCAGCTGCACTTGA 446
Db      426 ATGCGTGAAGATTCGCTACCGCGGCATGTGAACCAAGAAAAGTTGTGCGGACTTGT 485
Qy      447 GCGGTATGCAAGGTTGCTTTTGAACGTTTGGGGAACGATCCAGATCGATCAACCAT 506
Db      486 CAATTACGCTCGCTTGTCTTCCAGAGACTCGCGCATCGCTCAAGCATTTGATCACGTT 545
Qy      507 CAACGAMCCCTGATTCAGGCACTATATGATATGCCAGCCGACCAACGCCCGGCGAG 566
Db      546 CAACGAACCTGTGTCTATACGCTGCTGATATGACGCGGCGCTCCACCCACCGGCGC 605
Qy      567 GAGCAGATTAAACAGATCTCCACCGGGAACAATGCCATGAGCGCTGCTGCTG 626
Db      606 GTCCTTATTCGCGGACCGCAACGAGAGGCGACTGCTCCACCGACCATTCATCTTGC 665
Qy      627 AAAGCCCATATCATGAGCCATGCGCCGCGCTGCGCTTACAGCAGGACCTTTCGCC 686
Db      666 GCACACGAGACTGTGCTCCACCGGACAGCTGTCCCTCTTACAGCAAGATTCAGCC 725
Qy      687 CTGCAAAAGGCGCAATCGGCACTTCGCTCAACGCGCACTATATGAGCCCTGGGACAG 746
Db      726 CCAATGACAAAGGACATCGGCATCAAGCTTCAAGGGAATGGTCCAGCATGGAGATGA 785
Qy      747 CAATGAGCTTCGGGACAAAGAGCTGCTGAGCGCAAGATGGAATTTTCACTTGGCTGTT 806
Db      786 AGCGGACCCCTGGATCAGACCGGCGCTGAGCGCGCGGAGTTGCAAGATTCGCTGTT 845
Qy      807 TGCCAATCCCATCTTCTTGAAGAAGACTATCCAGAGAGCATGAAGAAGAGCTGGGCGA 866
Db      846 CGCGGACCCCTCTTACAGACGAGGTGCTATCCGCTCTCATGAGAGCCAGCTGGGTGA 905
Qy      867 GAGGCTTCCAGCCCTCACTCCCGGGAATTTCCTCATCTCAATGCGGAGAGACCGACTT 926
Db      906 CCGTGTGCGCAAGTTTCAAGCCGAGGAGT---CAAGCTGTGTGCTTGGAGCTCCGAGTT 962
Qy      927 CTACGCGATTAATTAATCAACATCCAGTTGCGCGGCACTGACGATCCCTGCCCGA 986
Db      963 CTATGCGATTAATCTGTAACAGACTTCTTCTGTCACAGACCAAGACACGCGGGAAT 1022
Qy      987 GAGGACTATCTGGGCGCATTCATGAGCAACAGAGATTAAGAGCGGACGCCGCTTGG 1046
Db      1023 TAACGACCAAGAGGGAATTAAGATTCATGATTTCAACAAACATGCACTTCTCGAGG 1082
Qy      1047 CGAGGAGCGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1106
Db      1083 CGAAGAGAGGACACCGAGTGGCTGGAGCTGCCCTGGGGCTTCCGAAAGCTTCTGAA 1142
Qy      1107 CCGGCTGACGCGCTGTAACGCAAGCCCATCTATCATACGAGAGAGGATCCCTGCC 1166
Db      1143 TTGATCTGTGCTCTGCTGATGAGATGCTCATCTATGTAACGAGAAAT---GACAGACG 1199
Qy      1167 TGAAGGAGAACATGAGCTGCGAGAGGCGCTCAAGACCCCTTCGCAATCGGTAAT 1226
Db      1200 CAAGGAGGAGACAGCCCTTCAACCAATGTTCTCAAGACCAATTTGCAACAGGTTCT 1259
Qy      1227 TGAATCGCATTTGACT---CGATTTCAAGGCCATTAACGAGACGCGCTGCTGCA 1283
Db      1260 TGAAGGATATGTTGGGTGGGCACTGCTCGTGGCTCAAGAGAGCGGATTAATCCG 1319
Qy      1284 GGGGTACTTTGGCGGCTTCTGCAATCTGG 1318
Db      1320 CTCCTATTTGCTGCAATTCACGATTAACCTGG 1354

```

RESULT 10
 US-09-417-507-19268
 ; Sequence 19268, Application US/09417507
 ; GENERAL INFORMATION:
 ; APPLICANT: KEITH G. WEINSTOCK ET AL.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
 ; TITLE OF INVENTION: FUNGICIDUS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: PAT99-10
 ; CURRENT APPLICATION NUMBER: US/09/417,507
 ; CURRENT FILING DATE: 1999-10-14
 ; SEQ ID NO: 19268
 ; LENGTH: 741
 ; TYPE: DNA
 ; ORGANISM: A. fumigatus
 US-09-417-507-19268

Query Match 23.0%; Score 334.4; DB 20; Length 741;
 Best Local Similarity 67.3%; Pred. No. 4,2e-66;
 Matches 470; Conservative 1; Mismatches 227; Indels 0; Gaps 0;

```

Qy      621 CGCTGAAAGGCCCAATCATGAGCAATGCGCGCGCGCTGAGCGCTTACAGCGGACTT 680
Db      6 CGCTGTGCGCCCATCCCGTGTGAATACCCGCGCGCTGCTGTACAAAGAGATT 65
Qy      681 TGCGCCCTGCAAAAGGCGAGATCGGCATCTGCTCAACGCGCACTATGAGCCCTG 740
Db      66 CCGTCTGTCCACAGAGGCAAGTATGAGATCTGTTGAACGAGACTATATAGCTTGG 125
Qy      741 GGAACCAATGAGCTTCGGGACAAAGAGGCTGCTGAGCGACGATGAATTTCAATGG 800
Db      126 GAATGCCAAGAGAGAGCGGATACGCTGTGAGAGCGGTGCGATGGAATTTCAATGG 185
Qy      801 CTGCTTTGCCAATCCCATCTTCTTGAAGAAGACTATCCAGAGCATGAAGAAGCACT 860
Db      186 CTGCTTCGCAATCTCTGCTTCTGCGACAGACTACCAAGCATGATGGAAGCAACT 245
Qy      861 GGGCGAGGCTTCCAGCCCTCACTCCGCGCACTTTCGCATCTCAATCCGAGAGAC 920
Db      246 TGAAGCTGCTACCAAGATTTCCTCCCTGTGATTTGACCTCTTCGCGAGCCGAGAG 305
Qy      921 CGACTTCTACGCGATGAATTAATCAACATCCAGTTGCGCGGCACTAGACGATCCGT 980
Db      306 CGATTTCTAGGAGATGAATCAATATCATCCAGTTTGTGCTGCGCATCGGATCAACAGC 365
Qy      981 CCCCAGAGCGACTATCTCGGCGCATTCATGAGCACAGAGAAATTAAGAGCGGAGCC 1040
Db      366 ATCCGAACAGATTAATTTGGCAATGATGAGTTAGATCAAGAGAAAGTAAAGGGAAC 425
Qy      1041 CGTTGCGAGGAGAGGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1100
Db      426 AGTTGGGAACCAAGGTATTCATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 485
Qy      1101 TCTGCGCCGGGTGTAAGCTGTACGCAAGCCCATCTACATCAACGAGAGATGCC 1160
Db      486 CTTCAAGAGATTATTCGCTGTACGGAACCAATTTTCATCAACGAGAGATGCTC 545
Qy      1161 GTGCGCTGAGAGAGAAACATGAGCTGCGAGAGGCGCTCAACGAGACCCCTTCG 1220
Db      546 ATGCTCTGAGAGAGATCGGATGACATGCGAGGAGTCAATCAAGATGATGAATCCG 605
Qy      1221 STACTTGAATCCCATTTGATCTGATTTCCAGGCGCATTAACGAGAGCGGCTGCT 1280
Db      606 GTATTTGAGATCAATCAAGGCTGTGGGCTGTCTGTCAACCAAGCGGCTGATAT 665
Qy      1281 CAAGGATACTTTGCGTGGCGTTGCTGCAATCTGG 1318
Db      666 CAGAGCTACTTTGCAATGCTGTGCTGATTAATTTAG 703

```

RESULT 11
 US-09-404-520-15403
 ; Sequence 15403, Application US/09404520
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Ghodsel, Azila
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: McIninch, James

```

; APPLICANT: Timberlake, William E.
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: Emericella nidulans Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15488)A
; CURRENT APPLICATION NUMBER: US/09/404,520
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 44345
; SEQ ID NO 15403
; LENGTH: 2918
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-09-404-520-15403
```

Query Match 21.9%; Score 318.6; DB 20; Length 2918;

Best Local Similarity 53.8%; Pred. No. 2,1e-62; Matches 701; Conservative 1; Mismatches 595; Indels 6; Gaps 2;

```

QY 16 GCTCTGCCCAAGCATTTGATGGGGCTTCGCAACGGCCGCTTACAGATCGAAGCGCC 75
DB 1147 GCTTGGCCCAATGACTTCTCCATGATATGCAACGGCGCGCAGGTTGAAGAGCG 1206
QY 76 GTCAAGAAGGAGCGCGCGCCGTCATCTGGGCAAGTACTGCACTGCACTGAGCGCATC 135
DB 1207 TGAACAAGATGAGCAAGGCTCTTGATCTGGGCAACATTCGGGCAATCTCAAGC 1266
QY 136 CGCAACCAAGCGCGCAAGCGGATGCTGGCTTGCATCTACCAACCGCTAGATGAGAC 195
DB 1267 GTCAAGAAGCAATGAGCAATGCGGATGACGCTGAGGTTCTACGACTTTTACCGGA 1326
QY 196 TTTGATCTCTTGACCAATGACGCGCAAGGCTTACGCTTCTCTGCTGCTGCGG 255
DB 1327 GTTGCCCTGATGAATGCTGATGCTGATGCTCAACGCTTATCTGCTGCTGCGG 1386
QY 256 ATCAATCCCTCGCGCGGAGGCTGATCCGTCACGAGGAGGAATGACTTTTACAGC 315
DB 1387 ATATCCCACTCGAGGCGCGGACGACCCGCTCATGAGCAAGGATCAAGTACTAC 1446
QY 316 AAATGATGACGCGCTTGTAGGCGGGATCAACGCTTGGTGTACTTTTACCACTG 375
DB 1447 GACCTGATGAGAGGACTACTCAATATGAGATTCAGGCTTCTGTAACCTTTTCACT 1506
QY 376 GATCTGCTCGGCGCTTCAAGATGCTGATGAGGCTGCTGCAACGTTGAAGAGTCC 435
DB 1507 GATGTCGCGAGGCTTGTGAAGATCGGATCGGGGCAATGCTGAACCGAGAGGTT 1566
QY 436 CTGACTTTGACCGGATGACGAGGTTGCTTGAACGTTTGGGAGCCGAGTCCAGAC 495
DB 1567 CTTGACTTTGCTCCGCTAGCGCGCTGTCTGTTCAACGCTTGAAGGCTTCCGCTAC 1626
QY 496 TGGATCAACATCAACGACCTGATTCAGGCACTATGATATGCAACCGAGCAAC 555
DB 1627 TGGATTCATTCATCAATGAACTGATGCTGATCTGCTGAGGTTTGGGAGGTTT 1686
QY 556 GCGCGCGGAGAGGAGCATTAACAAGATCTCAACCGAGGCAACCTGCACTGAGCC 615
DB 1687 GCTCTGACGCGTATGCTTTCGCAACTCAATGAAGAGGTTGCTTCTGCAAGAAC 1746
QY 616 TGGCTGCTGGAAGAGCCAGATCATGAGCCATCCGCGCGCTGAGCGCTTACAGAG 675
DB 1747 TTTATCGTCGCGGCAACAAACTGTGACCCAGCGCATGTTTCCAGCTTTTACCGAG 1806
QY 676 GACTTTGGCCCTCGCAAAAGGCGCAATGCGCATCTGCTCAACGCGCATCTATGAG 735
DB 1807 GTGTTCCAACCGGAGCAAGAAAGCAATTTGGCATCACTCCATGCACTGCTGCA 1866
QY 736 CCCTGGGACAGCATGAGCTCGGGAGCAAGAGGCTCTGAGCGACAGATGGAATTTTAC 795
DB 1867 CTTTGGAGTGAAGCATGCGGAGGACAGAGCAACAGAGGCGCAAGGAAATTTGAG 1926
QY 796 ATTGCTGCTTGGCAATCCCATCTTCTGGAAGAGACTATCCAGAGAGATGAAGAG 855
DB 1927 ATGCTTGTGTTCCGGAGCCCTTATACAGAGAGGCACTACCAAGCTTCAATGCGG 1986
```

```

QY 856 CAGCTGGGAGAGGAGGCTTCCAGGCTTCACTCCCGGAGCTTTGCAATCTCAATGCCGA 915
DB 1987 CAATTCGGAGACGCTTCAACCGCTTCTACTCCGAGAGT---CAAACTGGTACTAGGA 2043
QY 916 GAGACCACTTCTTACCGCATGAATTTACTACATATCCAGTTTCCGCGCCACTAGACGT 975
DB 2044 AGTTCAATATTTTACGATATGAATGACTGATACAGCACTTTTTCGAGAGCAAGAGATAG 2103
QY 976 CCGCTCCCGAGAGAGACTATCTCGGCGCATTCATGAGACCAAGAGAAATTAAGAGCGC 1035
DB 2104 CTTCCAGATATCATATGACCAAAAGCAATGATCATTTGTTACCATTCMAACAGCAAGGC 2163
QY 1036 AGCCCGCTTGGAGAGAGAGCGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1095
DB 2164 GTCTCTGAGAGCGCAAGATGCTGATGAGCTTGGCTTGGCTGAGACCGGACCATGATGAGCG 2223
QY 1096 AAGCATCTGCGCGGCTGTAAGGCTGTAAGGAGGCACTTCAATCAACGAGAAAGCA 1155
DB 2224 AAGCTGCTCAATTTGATTTGAACCGATATCATGCTGCTATATATGTCACGAAATG-- 2281
QY 1156 TGCCTGCGCTGAGAGAGAGAACTGACGTGAGAGAGGCGTCAACGACCTTCCG 1215
DB 2282 -GACGACAGCAAAAGAGAGACAGCTCCACACAGAGGACTCATTCGACATTTCCGT 2340
QY 1216 ATCCGTAATTTGACTGCTGCACTTGAATTCGCAAGGCTTACCCAGAGGAGCTC 1275
DB 2341 ATGCAATTTCTGAAAGGCTTACGTTGATGATTTGGCCCGTCTCTCAAGAGATGAGATG 2400
QY 1276 GTGTCMAAGGGGTAATTTGCTGCGTGGCGTGTCTGATTAATTTGG 1318
DB 2401 GATATCGGCTTACTTGTGCTGAGCTTCAACGCAACATGGG 2443
```

RESULT 12

US-10-369-493-37092

; Sequence 37092, Application US/10369493

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 37092

; LENGTH: 1470

; TYPE: DNA

; ORGANISM: Aspergillus nidulans

US-10-369-493-37092

Query Match 21.8%; Score 317.8; DB 50; Length 1470;

Best Local Similarity 53.8%; Pred. No. 3e-62; Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;

```

QY 20 TGGCCAAAGCATTTGAAATGGGCTTTCGCAACGGCGCTTACAGATCGAAGCGCGTCA 79
DB 2 TGGCAATGATCTTTCATGATATGATGACAGCGCGCGGCGAGGTTGAAGAGGCTGGA 61
QY 80 AAGAGGTGGCGCGCGCGCTTCTCATTTGGGACAGGATCTGCTCACTGAGGCAATGCGCA 139
DB 62 ACAAGATGAGCAAGGATCTTTCATCTGAGACATTTGCGGATATCTCCAGCAAAAGTCA 121
QY 140 CCAAGCGCCCAAGCGGATGCTGCTTGGATCACTACACCGCTTACATGAGAGCTTTG 199
DB 122 AAGACATATGCAATGAGATGAGCGCTGAGGTTTCTACGACTTTTACCGCAAGATGTTG 181
QY 200 ATCTTTGACCAAGTACGCGGCAAGGCGCTTACCGCTTCTTGTGCTGCTGCGGATCA 259
```

Db 182 CCTGATGAGTGTATGATGTCACGCTTATGATTCGTATCTCGATCGGGCATTA 241
 Qy 260 TTCCCTCGGGCGCAGCTGATCCGCTCAACGAGAGGAAATTGATTTACAGCAAC 319
 Db 242 TCCCACTCGAGGCGCGACGACCCGGTCAATGAGCAAGGAACTCAATCTACAGACC 301
 Qy 320 TGAATTGACGCGCTGTGAGGCGGGGTATCAAGCCTTGAGTGTGATCTTACCACTGGATC 379
 Db 302 TGGTATACAGCTACTCAATATGGGATTAAGCCATTCGTAACCTTTTCACTGGATG 361
 Qy 380 TGGCTGAGGCGCTTCAAGATCGCTATGAGAGCTGGCTCAACGTGAGAGAGTCCAGTGG 439
 Db 362 TTCCGAGGCTTTGAGAGATCGGATAGGGGCAATGTGAACAGAGAGGTTCACTTCCTG 421
 Qy 440 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAGAGCTTTGGGAGCCAGTCCAGAACTGGA 499
 Db 422 ACTTTGTCGCTAGCGCGGTGTCTGTTTCAAGCGCTTGAAGGCGCAAGAGTCCGTCACTGA 481
 Qy 500 TCACCATCAACGAMCCCTGGAATTCAAGGCATCTATGGAATGCAACCGGACGCAACGCC 559
 Db 482 TTACATTTCAATGAACCTGATGTGTATCTGCTGCGAGGTTATGCGGACGATGTTCAATGCTC 541
 Qy 560 CGGCGAGAGACACATTTAACAGCACTCAACGAGGCAACACTGCCACTGAGCGCTGAC 619
 Db 542 CTGACGATCATGCTTTTCCGGAATCAATGAGAGGTTGATTCCTGACAGAACCTTTCA 601
 Qy 620 TCGCTGAGAAAGGCCAGATATAGACCTCCGCGCGCTGGCGCTTACAGAGGACT 679
 Db 602 TCTGTGCGCCACACAAAATGTGTGATCCACGCGCATGTTTCCAAAGCTTTTACGAGAGTGT 661
 Qy 680 TTGCGCCCTTGCAAAAGGGCCAGATCGGCATCTCGCTCAACGGGACTTACTATGAGCCCT 729
 Db 662 TCCAACTGCGACGAAAGAAACAATTTGATCACTTCACTGACCTGATGCTGGAACCTT 721
 Qy 740 GGGACAGCAATGAGCTCGGAGACAGAGGCTGTGACGACGAGTGAATTTTCAATTTG 799
 Db 722 GGGATGAGACGATTCGCGGGGACAGAGACAGAGAACGGGCGAGGAAATTGAGATTCG 781
 Qy 800 GCTGTGTTTGGCAATTCCTATCTTTTGAAGAAGACTTCCAGAGACATGAAGAGAC 859
 Db 782 CTGTGTTCGCGACCCCTTATACAGACAGGCACTTACCAAGCTCAATCGGCTTCAC 841
 Qy 860 TGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTGTGCATCTCAATCGCGGAGAGA 919
 Db 842 TCGGGGACGCTTACCGCGCTTCACTCCGAGAGT---CAAACTGTGATCTAGAAAGTT 898
 Qy 920 CCGACTTCTACGCGCATGATTTACTACATCCAGTTGCGCGGCACTTGAACGATCCG 979
 Db 899 CAGAAATTTTACGATATGAACTCGTACAGACCTTTTTCGTGACAGACAAAGATACGCTC 958
 Qy 980 TCCCGACAGACGATCTATCTGCGGCGCATTCATGACACAGAGAAATTAAGACGCGAC 1039
 Db 959 CAGATATCATATGACCAAAAGGCAATGTCTATTTTCAATACCAACAGAAAGGCTCTT 1018
 Qy 1040 CGGTGCGAGAGAGAGGCGCTGCGCTGCGGCTGCTGCTCCGAGACATTTCCGGAAGC 1099
 Db 1019 CTGAGAGCGAGAGTCCGATACGCTTGTGCTGCGAGCGGACCCCACTGGAATGGGCAAGC 1078
 Qy 1100 ATCTGCGCGCGGTGTATCGGCTGTATCGGCAAGCCCATCTACATCAACGAGAACGATGCC 1159
 Db 1079 TGTCTCAATTTGATTTGGAACCGATATATGCTGCTATATATGTCACGGAATAATG---GCA 1135
 Qy 1160 CGTGCGCTGAGAGAGAGACATGACGTGCGAGAGAGCGCTCAACGACCCCTTCCGATCC 1219
 Db 1136 CGACAGCAAAAGAGAGAGACGCTCCACACAGAGGATATCATCTGACACTTTCCTGATGTC 1195
 Qy 1220 GATACCTTGAATCTGCACTTGGATCGATTTCAAGGCTATTAACCAAGAGCGGCTGCTG 1279
 Db 1196 GATTTCTTCAAGGCTACGTTGTGAGATTTGGCCGCTGCTTCAAGAGAGATGAGATGATA 1255
 Qy 1280 TCAAGGAGTACTTTGCGTGGCGTGTGCTGATTAACCTTGG 1318

Db 1256 TCCGATCTTACTTTGCTGAGCCTTCACGCAACATGGG 1294
 RESULT 13
 US-60-360-039-37092
 ; Sequence 37092, Application US/60360039
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Chen, Xianfeng
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)A
 ; CURRENT APPLICATION NUMBER: US/60/360, 039
 ; CURRENT FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 37092
 ; LENGTH: 1470
 ; TYPE: DNA
 ; ORGANISM: *Aspergillus nidulans*
 US-60-360-039-37092
 Query Match 21.8%; Score 317.8; DB 93; Length 1470;
 Best local Similarity 53.8%; Pred. No. 3e-62;
 Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;
 Qy 20 TGGCCAAACGATTTGAATGGGCTTGGCAACGCGCGCTTACCAATGAGAGCGCGCTCA 79
 Db 2 TGGCGAATGATCTTTCATGATATGCAACGCGCGCGAGGTTGAAGAGCGCTGA 61
 Qy 80 AAGAAGTGGCGCGCGCGCGCTTCATCTGAGACAGTATGCGACCTGAGACCATCGGCA 129
 Db 62 ACAAAGATGCAAGGATCTTGTGATCTGGACCATTCGGGCAATTCACGCAAAAGTCA 121
 Qy 140 CCAACGCGCGCAACGCGAGATGTGCTTGCATCTACCAACCGCTACGATAGAGACTTTG 199
 Db 122 AAGACATAGCAATGGGATGAGCTGTGAGCTTACAGACTTTTACCGGAAAGATGTTG 181
 Qy 200 ATCTCTTGAACAGTACGCGGCAAGGCTTACCGCTTCTCTTGTGCTGCTGCGGATCA 259
 Db 182 CCTGATGAAGTGTGTGATGCTCAACGCTTATGATTCGTATCTGCTATCTGCTGCGCATTA 241
 Qy 260 TTCCCTCGCGCGCAGGCTGATCCCGTCAACAGAGAGGAAATGATTTTACAGCAAC 319
 Db 242 TCCCACTCGAGGCGCGCAACGACCCGCTCAATGAGAGAGGATCACTACAGAGACC 301
 Qy 320 TGAATTGACGCGCTGTGAGGCGGGGTATCAAGCCTTGAGTGTGATCTTGTACCACTGGATC 379
 Db 302 TGGTATACAGCTACTCAATATGGGATTAAGCCATTCGTAACCTTTTCACTGGATG 361
 Qy 380 TGGCTGAGGCGCTTCAAGATCGCTATGAGAGCTGGCTCAACGTGAGAGAGTCCAGTGG 439
 Db 362 TTCCGAGGCTTTGAGAGATCGGATAGGGGCAATGTGAACAGAGAGGTTCACTTCCTG 421
 Qy 440 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAGAGCTTTGGGAGCCAGTCCAGAACTGGA 499
 Db 422 ACTTTGTCGCTAGCGCGGTGTCTGTTTCAAGCGCTTGAAGGCGCAAGAGTCCGTCACTGA 481
 Qy 500 TCACCATCAACGAMCCCTGGAATTCAAGGCATCTATGGAATGCAACCGGACGCAACGCC 559
 Db 482 TTACATTTCAATGAACCTGATGTGTATCTGCTGCGAGGTTATGCGGACGATGTTCAATGCTC 541
 Qy 560 CGGCGAGAGACACATTTAACAGCACTCAACGAGGCAACACTGCCACTGAGCGCTGAC 619
 Db 542 CTGACGATCATGCTTTTCCGGAATCAATGAGAGGTTGATTCCTGACAGAACCTTTCA 601
 Qy 620 TCGCTGAGAAAGGCCAGATATAGACCTCCGCGCGCTGGCGCTTACAGAGGACT 679
 Db 602 TCTGTGCGCCACACAAAATGTGTGATCCACGCGCATGTTTCCAAAGCTTTTACGAGAGTGT 661
 Qy 680 TTGCGCCCTTGCAAAAGGGCCAGATCGGCATCTGCTCAACGCGGCACTATGAGCCCT 729

```

Db      662  TCCAACTCCGACGAGAAAGAAACAATTGGCATACCTTCATGCGAATCGGTCCGAACCTT 721
Qy      740  GAGACAGCAATAGACCTCGGAGACAAGAGGCTGTGAGCGACGATGGAATTTTCACATTG 799
Db      722  GGGATGAAAGCGATCCGCGGAGACGAGAGACGAGAACGAGGAATTCGAGATCG 781
Qy      800  GCTGGTTTGGCAATCCATCTTTTGAAGAAGACTATCCAGAGAGCATGAAGAAGACG 859
Db      782  CTGGTTTCCGAGACCTCTTATACAAACAGAGCATACCCAGCTCAATGGGGCTCAAC 841
Qy      860  TGGGCGAGAGGCTTCAGACCTCCTCCTCGGAGACTTTGCAATCTCAATGCCGAGAGA 919
Db      842  TCGGGAGACCTCTACCCGCTTTCCTCCAGAGAGT---CAAACTGTACTAGGAATG 898
Qy      920  CCGACTTCTACGCGCATGATTTACTACAATCCAGTTGCGCGCCACCTTAGACGCTCCG 979
Db      899  CAGAAATTTTACGGTATGAATCTGACACGACCTTTTCTGTGACACAGAGATACGCTTC 958
Qy      980  TCCCCGAGACGAGACTATCTCGCGCCCATCCATGAGCACAGGAAATTAAGACGCGAGCC 1039
Db      959  CAGATATCATATGACACAAAGGCAATGTATTTTCAAGATCAAAACGACAAAGGCTCT 1018
Qy      1040  CCGTTGGCGAGAGAGCGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1099
Db      1019  CTCGAGGCGAGAGATCCGATACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1078
Qy      1100  ATCTCGCCCGGGTGTACGCGCTGTACGCGCAAGCCATCTACATACCGAGACGATGCC 1159
Db      1079  TGCTCAATTTGGATTGGAAACGATATCATGTGCTATATATGTACACGGAATATG---GCA 1135
Qy      1160  CGTCCCTGAGAGAGAGAACTAGACGTGCGAGAGGCGGTCAAGCCCTTCGCGATCC 1219
Db      1136  CGACAGCAAAAGAGAGACAGCTCCACACACAGAGATCTCATGACCTTTCCGATGC 1195
Qy      1220  GGTACTTTGACTCCGACTTGGACTCGATTTCCAAAGCCATTAACCAAGACGCGCTGTGC 1279
Db      1196  GATCTTCGAAAGGCTAGGTTGTGTGATTTGCGCCGCTGCTGCTCAAAAGAGATGAGATG 1255
Qy      1280  TCAAGGGGACTTTTGGCGTGGCGCTTGTCTCAATTAATTGG 1318
Db      1256  TCCGCTTACTTGTCTGCTGACCTTCAACCGAACCTGGG 1294

```

```

RESULT 14
US-60-138-103-7477
; Sequence 7477, Application US/60138103
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Timberlake, William E.
; TITLE OF INVENTION: ASPERGILLUS NIDULANS GENOME SEQUENCE AND USES THEREOF
; FILE REFERENCE: 38-101154851D
; CURRENT APPLICATION NUMBER: US/60/138,103
; NUMBER OF SEQ ID NOS: 28006
; SEQ ID NO 7477
; LENGTH: 2914
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-60-138-103-7477

```

```

Query Match      21.1% Score 307.6; DB 69; Length 2914;
Best Local Similarity 53.8% Pred. No. 7,1e-60;
Matches 701; Conservative 1; Mismatches 595; Indels 7; Gaps 3;

```

```

Qy      16  GCTCTGCCCAAGCATTTGAATGGGCTTCCGAACGCGCCCTACCAAGTCCAGAGCGCC 75
Db      1147  GCTTGGCCCAATGACTTTCCATGATATGCAAGCGCGCGCGAGGTTGAAGAGCG 1206
Qy      76  GTCAAGAAGTGGCGCGCGCGCTTCATCTGGGACAGTACTGCACTGAGGCAATCG 135
Db      1207  TGGAAACAAGATGCAAGAGGCTCTTCGATCTGGGACACATTCGGGCTATCTCCAGGCAA 1266

```

```

Qy      136  CGACCAACGCGCGCAACGCGATGTGCTTGGCATCACTACCAACCGCTACGATGAGAC 195
Db      1267  GTCAAGACATATGATGCGATGAGACGCTGTGAGGTTCTACAGACTTTTACCGCAAGAT 1326
Qy      196  TTTGATCTCTTGAACCAATGACGCGCAAAAGCTTACCGCTTCTCTTGTGTGTGGCG 255
Db      1327  GTTGCCCTGATGAAGTGTATGTGTCAACGCTTATCATGATTTCTCGCTATCTGTGTGCG 1386
Qy      256  ATCATTCCTCGCGCGCGAGCTGTATCCGTCACAGAGAGGGAATTGATTTTACAGC 315
Db      1387  ATATCCCACTGGAAGCGCGCGACCGGTCACAGACCGGTCAATGAGAGAGATCAATGATCA 1446
Qy      316  AAATGATTTGACCGCTGTGAGCGGAGTATACGCTTGGTGAATTTGATCACTGG 375
Db      1447  GACTGTGATGAGAGTACTATATATGAGATTAACGCAATTCGTAACCTCTTCACTGG 1506
Qy      376  GATCTGCTTACGCGCTTACAGATTCGCTATGAGAGCTGGCTCAACGTGGAAGGTCCAG 435
Db      1507  GATGTTCCGAGGCTTTGAGAGATCGGTACGCGGGGATGCTGAACACAGAGAGGTTCATT 1566
Qy      436  CTGACCTTTGAGGAGTATGCGAGGTTGTGCTTGAACGTTTGGGGAACGAGTCCAGAAC 495
Db      1567  CTGACTTTGTCCGCTACGCGCGCTGTCTGTTTGAACGCTTAGGCGCAAGGTCCGCTCAC 1626
Qy      496  TGAATCACATCAACAGMCCCTGG-ATTCAAGCATCTATATGATATGCAACGCGAGCAA 554
Db      1627  TGGATTTACATTAATGAAACATGTGTGTATCTGCTGCGAGGTTATGCGCAGGTTTCA 1686
Qy      555  CGCCCGGAGAGAGACATTTAAACAAGCACTCAACGAGGCAACATCTGCCCTGAGCC 614
Db      1687  TGCTCTGACGCGTATCGTTTGTGCGAACTCAATGAAGAGGGTATTCCTGACAGAAC 1746
Qy      615  GTGCTGCTGAGAAAGCCCAAGATCATGAGCAATGCCGCGCGCTGCTACAGCAG 674
Db      1747  GTTCATGTGCGCCACACAAATGCTGACCCAGGCAATGTTTCCAGCTTTACGAGAG 1806
Qy      675  GAACTTTGCGCCCTCGCAAAAGGCGCAGATCGGATCTGCTCAACGCGCACTATGA 734
Db      1807  GGTGTTCCAAACCGAGAGAAAGAACATTTGGCATCACTCCATGAGCAATCTGTGCGA 1866
Qy      735  GCCCTGGAAGACATATAGCTTGGGACAGAGAGCTGTGAGAGACGATGGAATTTCA 794
Db      1867  ACCTTGGGATGAAGACATCCGCGGACAGAAAGACAGACCGGCGCCAGGGAATTCGA 1926
Qy      795  CATTTGCTGTGTTGCAATCCCATCTTTTGAAGAAAGCTATTCAGAGAGCATGAAGAA 854
Db      1927  GATGCTTGTGTTCCCGAACCCCTTATCAAGACAGGAGCTTACCCAGCTCAATGCGGGC 1986
Qy      855  GCACTGTGGGAGAGAGGCTTCCAGCCCTCACTCCCGCGAATTTGCCATCTCAATGCCG 914
Db      1987  TCAACTGGGAGACGCTTACCGCGTTTCACTCCGAGAGAT---CAAACTGTGATCAAG 2043
Qy      915  AGAACCGCATTTCAAGGCAATTAATTAACAATCCAGTTGCGCGGCACTAGACGG 974
Db      2044  AAGTTCAAGAAATTTTACGATATGAATCTGTACACACACTTTTGTGTGAGACAAAGATAC 2103
Qy      975  TCCGCTCCCGAGACGAGATATCTCGGCGCATTCATGAGACCAAGAGATTAAGACGG 1034
Db      2104  GCTTCAAGATATCAATATCAACAAAGGCAATGATGTTGTTCAAGATCAAAACAGCAAGG 2163
Qy      1035  CAGCCCGTTGGAGAGAGAGCGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1094
Db      2164  CGTCTCTGAGGCGAAGAGTCCGATACGCTTGTGCTGCGAGCGACCACTGATGAGCG 2223
Qy      1095  GAACATCTGCGCGCGGTATACGCGCTGTACGGAACCCATCAATCAACCGAGAACGG 1154
Db      2224  CAAGCTCTCAATTAATTTGAACCGATATCATGTGCTTATATGTCAAGAAAATG- 2282
Qy      1155  ATGCGCTGCTTGAAGAGAACATGACGTGCGAGAGCGCTCAACGACCCCTTCCG 1214
Db      2283  --GACCAACAGCAAAAGAGAGACAGCTCCACACAGAGGTACTATCACTTCCG 2340
Qy      1215  CATCCGTAATTTGATCTGCACTTGAAGCTGATTTCCAGGCGCATTCACAGAGCGGCT 1274

```

Db 2341 TATGCGATTCTCGAAGCTACGTGGTGGATTGGCCCGTGTCTCAAGAGATGAGT 2400
 Qy 1275 CGTCGTAAGGGGACTTTTCCGTGGGGCTGCTCGAATCTTG 1318
 Db 2401 GGATATCCGGCTTACTTGTGCTGAGACTTACCGACACTGG 2444

RESULT 15

US-09-404-520-19929
 ; Sequence 19929; Application US/09404520
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Ghodest, Azila
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: McIninch, James
 ; APPLICANT: Timberlake, William E.
 ; TITLE OF INVENTION: Emerice11a nidulans Genome Sequence and Uses Thereof
 ; FILE REFERENCE: 38-10(15498)A
 ; CURRENT APPLICATION NUMBER: US/09/404,520
 ; NUMBER OF SEQ ID NOS: 44345
 ; SEQ ID NO 19929
 ; LENGTH: 1434
 ; TYPE: DNA
 ; ORGANISM: Aspergillus nidulans
 ; US-09-404-520-19929

Query Match 20.2%; Score 294.6; DB 20; Length 1434;
 Best Local Similarity 53.7%; Pred. No. 6,2e-57;
 Matches 676; Conservative 1; Mismatches 575; Indels 7; Gaps 3;

Qy 61 CAGATCGAAGCGCCCTGCAAGAAAGGTGGCGCGCCCGCTTCATCTGGGACAGTATCG 120
 Db 3 CAGGTGAAAGAGCGTGGAAAGAGATGCGAAGGCTCTTCGATCTGGGACACTTCGGG 62
 Qy 121 CACCTGAGCGCATGCGCCCAACGCGCGCCACGCGCATGCGCTTGCACTACAC 180
 Db 63 CATATCTCGAAGCAAGCTCAAGACATAGCAATGCGATACGCTTGAGAGTTCTAGAC 122
 Qy 181 CGGTACGATGAGACTTTGATCTTGTACCAAGTACGCGCAAGGCTTACCGTTCTCC 240
 Db 123 TTTTACGCGAAGATGTGCTCGATGATGATGATGATGATGATGATGATGATGAT 182
 Qy 241 TTGTCTGTGCTGCGGATCATTTCCCTTGGCGGAGCTGATCCCTGCAAGAGGGA 300
 Db 183 CTATCTGTGCTGCGGATCATTTCCCACTCGAGGCGCCGACGACCGGCTCAATGAGCAAGG 242
 Qy 301 ATGAGTTTACAGCAAACTGATTGAGCGCCTGTTGAGGGGGTATCAGCCCTTGGGTG 360
 Db 243 ATCAAGTACTTACAGAGCTGTGATGAGAGCTACTCAATATGAGGATTTACGCAATTCGTA 302
 Qy 361 ACTTTGTACCACTGGGATCTGCTCAGCGCTTACAGATCGCTATGAGGCTGCTCAC 420
 Db 303 ACCCTCTTACCTGGGATGTTCGCAAGGCTTTGGAAGATCGGTAAGGGGCGATGTAAC 362
 Qy 421 GTGGAAGAGGTCCAGCTGGACTTTGAGCGGTATGCAAGTTGCTTGAAGTTTGGG 480
 Db 363 CAGGAGAGGTTCACTCTGACTTGTGCTGACGCGCTGTCTGTTGAAACGCTAGGG 422
 Qy 481 GACCGAGTCCAGAACTGGATCACCATCAAGACCCCTGG--ATTGAGGCACTTATGATA 539
 Db 423 CCAAAGTCCGTGACTGAGATTTACATTTAATGAAACATGTGTGCTGCTGGCAGGTTA 482
 Qy 540 TGGCAACCGGAGCAAGCGCCCGGAGAGAGAGAGCATTTACAGCACTTCCAGCGGCAA 539
 Db 483 TGGGAGAGGTTCATGCTCTGAGAGGTGATGTTTGGGAATCAATGAGAGGAGTGA 542
 Qy 600 CACTGCACTGAGCGGTGCTCGCTGGAAGGCCAGATCATGAGCCATGCCCGCGCT 659
 Db 543 TTCTCGACAGAACCTTTCATGTCGGCCACACAAACTGATGATCCACGCGCATGTTTC 602

Qy 660 GGCCTCTACAGGAGGACTTTTCCCGCTTCGCAAAAGGCGCAGATCGCATCTCGCTCAA 719
 Db 603 CAAGCTTTACGAGAGGTGTTCCACCGCAGAGAAAGAAATTTGGATTCACCTCCCA 662
 Qy 720 CGGCGACTACTATGAGCCCTGGGACAGCAATGAGCTCGGAGACAGAGGCTGTAAGC 719
 Db 663 TGGCAACTGCTCGAAGCTTGGGATGAGACGATCCGCGGACCAAGAACAGCAGACCG 722
 Qy 780 ACGGATGAAATTTCACTTGGCGTGTGCTTGGCAATCCCATCTTCTTGAAGAGACTATCC 839
 Db 723 GCGCAGGAAATTCAGATGCTGTGTTCCCGACCTTATACAGACAGGCACTACCC 782
 Qy 840 AGAGACATGAAAGACAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGACTTTGC 839
 Db 783 AGCTCAATGCGGGCTCACTGGGAGCGGTACCGCGTTTCACTCCGAGAGT---C 839
 Qy 900 CATCTCAATGCGGAGAGACCGACTTTCAGGCAATGATTTACTACATCCAGTTGCG 959
 Db 840 AAAACTGTACTAGGAATTCAGAAATTTTACGGTATGAACTCGTACACGACCTTTTTCGT 839
 Qy 960 GCGCCACTAGAGGTCCTCCGCGAGACGACTATCTCGGCGCATCGATGAGACCA 1019
 Db 900 GCAGCAAGAGATACGCTTCAGATATCAATGACCAAGAGCATGTCTATGTTACGA 959
 Qy 1020 GGAAGATTAAGAGCGGAGCGCCGTTGGGAGAGAGCGGCTCGCTGCGCTGCTCTG 1079
 Db 960 TACAAACAGAAAGGCTCTCTGAGGCGGAGAGTCCGATACGCTTGGTGGAGCGG 1019
 Qy 1080 CCGGACATGTTCCGAAAGCATTCGCGCGGATGTACGGCTGTACGGAAGCCCATCTA 1139
 Db 1020 ACCCACTGATGCGGCAAGCTGCTCAATGATTTGAAACCGATATCATGTGCTATATA 1079
 Qy 1140 CATACCGAAGAGATGCCCCGCTGGAGAGAGAAATGACATGACGCGAGAGCGCT 1199
 Db 1080 TGTACGAAATG---GACGACAGCAAAAGAGAGACGCTCCACACCAAGGATCT 1136
 Qy 1200 CAAGCAACCTTCCGATCGGTAATCTTGTGACTGCACTTGGACTCGATTTTCAAGGCA 1259
 Db 1137 CATGACACTTTCGATGATGATTTTCGAAGGCTACGTTGTGATGTGCGCGCTGTGT 1196
 Qy 1260 TACCAAGAGCGGCTGCTCAAGGAGTACTTTGCGTGGCGTTGCTCGATTAATCTGG 1318
 Db 1197 CAAGAGAGATGAGTGAATATCCGATCTTAATTTGCTGGAACCTTACCGACAACTGGG 1255

Search completed: March 29, 2004, 02:51:02
 Job time : 5017.04 secs

This Page Blank (uspto)

LENGTH: 2016
TYPE: DNA
ORGANISM: Homo Sapien
US-10-152-372-253

Query Match 12.2%; Score 178.2; DB 6; Length 2016;
Best Local Similarity 54.5%; Pred. No. 2.3e-38;
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

```
15 AGCTGCGCCCAAGACTTTGATGGGCTTCCGCAACGGCCGCTACAGATGAAGGCGC 74
237 AACCTTCCCTTGTGGCTTCTCTGGGAGGCTGGGAGTTCTTGCTTACAGACGAGGCGGC 296
75 CGTCAAGAGGTGGCCGCGCCGCTCCATCTGGGACAGTACTCCACCTTGAGCANTC 134
297 CTGGGACGAGACGGGAAAGGGCTTAGCATCTGGGACGCTTTCAACACAGTGGGAAAGG 356
135 GCGCACCAAGC---CGCCACGCGGATGTGGCTTGCATCACTACCAACCGCTACGATGA 191
357 GAAAGTGTCTGGGAATGAGACGCGCAGATGTAGCTGTACGCGCTACTACAAGGTCCAGGA 416
192 GGACTTGTATCTTGAACCAAGTAGCGGCAAGGCTTACCGCTTCTCTGTGTGTC 251
417 GGAATCATCTTCTGTGAGGAACTGCACGTCAACCACTACGATTTCTCTGTGTGCGC 476
252 GCGGATCATTTCCCTCGGCGGAGCTGATCCGCTCAACGAGGAGGAAATTTGATTTTA 311
477 CCGGCTCTGCGCCACAGGCAATCCGAGCGGAGAGGTGAACAAAGGAAATCGAATTTTA 536
312 CAGCAAACTGATATGACGCGCTTGTAGGCGGGGATATCAACGCTTGGTGAATTTTGA 371
537 CAGTATCTTATGATGAGCTTCTTGAAGCAACATCACTCCATCTGACCTTGGACCA 596
372 CTGGGATCTGCTTCAAGGCTTCAAGATCGCTATGAGAGGCTTCAACGATGAAGGAT 431
597 CTGGGATCTGCTTCAAGGCTTCAAGATCGCTATGAGAGGCTTCAACGATGAAGGAT 656
432 CCAAGTGAATTTGAGCGGATATGCGAGGTTGTCTTGAACGTTTGGGAGCGAGTCCA 491
657 CAA---CTACTTCAGAGACTACGACCACTGTGCTTGAAGCCTTGGGAGCGGATGA 713
492 GAATGATTCACATCAACGACGACGCTTGAATTCAGGCTTATGATATGCGACCGGAG 551
714 GCATGATTCACATTCAGTATCTTGGGCAATGCGGCAAAAGGCTATGAGCGGCGCA 773
552 CAAGCGCCCGGCGAGGACGATTTAACAAGCACTCCACGAGGCAAACTGCGACATGA 611
774 CCAATGCGCGGGGCTGAAAGCTCCG-----GGACCGG 806
612 GCGGTGCTCGTGAAGAGCCAGATATGAGCCAGTCCGCGCGCTGCGCTCTACAG 671
807 CCTGTACAAAGGACACACATCATTTAAGGCCACGCAAACTGCGATTTCTTATA 866
672 CAGGGAATTTGGCCCTGCAAAAGGCGCAGATGCGATCTCGCTCAACGCGGACTACTA 731
867 CACCACTGCGGCGCAAGAGCAAGGCTGTGTGGAATTTCACTGAACTGTGACTGGG 926
732 TGAAGCTTGGGACGAGATGAGCTTGGGACAAAGAGGCTGCTGAGGACGAGATGAAT 791
927 GGAACCTGTGGAATTTGTAACCCCAAGGACCTAGAGGCTGCCAGAGATACCTACGTT 986
792 TCACATTTGCTGTGTTTGGCAATCCCATCT 820
987 CTGTGTGGGCTGTGTGCAACCCCATTT 1015
```

RESULT 2
PCT-US04-02242-95
Sequence 95; Application PC/TUS0402242
GENERAL INFORMATION:
APPLICANT: Diversa Corporation
APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN

TITLE OF INVENTION: THEM
FILE REFERENCE: 56446-20118.40
CURRENT APPLICATION NUMBER: PCT/US04/02242
CURRENT FILING DATE: 2004-01-26
PRIOR APPLICATION NUMBER: 60/442,794
PRIOR FILING DATE: 2003-01-24
NUMBER OF SEQ ID NOS: 224
SOFTWARE: PatentIn version 3.1
SEQ ID NO 95
LENGTH: 1374
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-95

Query Match 11.2%; Score 162.8; DB 1; Length 1374;
Best Local Similarity 60.0%; Pred. No. 2.8e-34;
Matches 329; Conservative 0; Mismatches 207; Indels 12; Gaps 3;

```
20 TGCCCAAGACTTTGAATGGGCTTGCACAGGCGGCTTACGATGGAAGCGCGTCA 79
38 TTCCCGCACCTTTCACCTGGGGCGTGCCACAGCGCTTACAGATCGAAGGCGCGCG 97
80 AAGAAGGTGGCGGCGCGCGCTTCCATCTGGGACAGTATGCGCACTGGAGCCATCGCGCA 139
98 CCAATGGCGGCGCGCGCGCTTCCATCTGGGATCTTCAAGCAGACAGAGGCAAGATCA 157
140 CCAAGCGGCGCAAGCGGATGTGGCTTGGATCACTACCAACCGCTACAGATGAGACTT 199
158 TCAGCGGCAAGATGAGGAGCTGGCTGCGACACTACCAACGCTATGCGAGAGCTGG 217
200 ATCTTGAAGAGTACGCGGCAAGGCTTCTCTGCTGTGTGGTGGATCA 259
218 AGCTGATGCGCAAGCTTGGGCTGAGACGCTTACCGCTTTCATGCTGTGGCGGCTCC 277
260 TTCCCTCGGCGGCGAGGCTGATCCGCTCAAGAGAGGGAATTTGATTTACAGAAAC 319
278 AGCCCAAGGTTTC-----GGCGCTGGAAGCAACAGGCTTTGATTTCTATGCGCGC 331
320 TGAATGAGCGCTTGTGAGGCGGGGATATCAAGCTTGGGATTTTGAACCTGAGATC 379
332 TGCTGACGCGCTTGGCGGCGCAAGGAGCTGACGCGCACTGACCTGTACCTGAGGAC 391
380 TGCTGAGGCGCTTCAAGATCGCTATGAGGAGGCTGCTCAACGTAAGAGATCAAGTGG 439
392 TGCGGAGGCTTGCAGAA---CGAGGGCGGCTGTGCTCAATGCGCCACTGTGTACC--- 445
440 ACTTGAAGCGGATATGAGAGTGTGCTTTGAACGTTTGGGAGCGAGTCCAGAACTGGA 499
446 ACTTGGCGGGATATGCGCGAGGTGAGGCGCGCGCTTGGCCACAAAGTCCAGATCG 505
500 TCACATTCACAGACGCTTGAATTCAGGCACTTATGATATGCAACGCGGACGCAAGCC 559
506 CCAAGCAGATGAGCGCTGTGTGATGCTGTGCGGCGACGAGCAGCGGCTGCGCGC 565
560 CCGGCGAG 567
566 CCGGCGATG 573
```

RESULT 3
US-10-767-701-7585
Sequence 7585; Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(53535)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29

```

; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 7585
; LENGTH: 805
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS93553_1
US-10-767-701-7585

```

```

Query Match      11.0%; Score 159.8; DB 6; Length 805;
Best Local Similarity 54.3%; Pred. No. 1.5e-33;
Matches 345; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

```

```

QY 20 TGCCCAACGACTTGAATGGGCTTCGCAACGCGCGCTACAGATCGAAGCGCGCTCA 79
DB 174 TCCCAAGAGACTTGTCTTGGGCAAGGCTCGCGGCTATCAATAGAGGCTCTACA 233
QY 80 AAGAAAGTGGCCGCGCGCTTCATCTGGGACAGTACTGCACTGCACTGAGCGCATCGCGCA 139
DB 234 ACGAAGAGGCAAGGCGCTACCATATGGGCAAGTTCATCATCTCCAGTAAAGATCT 293
QY 140 CCAAGCGCGCAAGCGGATGTGCTTGGCATCTACCAAGCTACAGATGAGACTTTG 199
DB 294 TGAACAACGATACCGGAGAGTACAGATGATGATGATGATGATGATGATGATGATGATG 353
QY 200 ATCTTGAACCAAGTACGCGGCAAGGCTACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTG 259
DB 354 AACTCTCAAGAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 413
QY 260 TTCCCTCGCGGCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 319
DB 414 TGCCAACGCTTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 473
QY 320 TGAATGACGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 379
DB 474 TCAATCAACGAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 533
QY 380 TGCTCAGCGGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 439
DB 534 CGCCCTCGCGGCTGAGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 590
QY 440 ACTTGAAGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 499
DB 591 ACTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 650
QY 500 TCACCATCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 559
DB 651 CCAAGTTCACGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 710
QY 560 CGGCGAGGAGCAGATTAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 619
DB 711 CGGCGCGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 770
QY 620 TCGCTGAAAGCGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 654
DB 771 CCAACACCGTGAAGCACAACATATCTCTCGCCAC 805

```

```

RESULT 4
US-10-767-471-414
; Sequence 414, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 6274
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(6274)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-3)
US-10-767-471-414

```

```

Query Match      10.4%; Score 151.6; DB 6; Length 6274;
Best Local Similarity 52.4%; Pred. No. 5.2e-31;
Matches 446; Conservative 1; Mismatches 371; Indels 33; Gaps 4;

```

```

QY 20 TGCCCAACGACTTGAATGGGCTTCGCAACGCGCGCTACAGATCGAAGCGCGCTCA 79
DB 4141 TTCTGAGGCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4200
QY 80 AAGAAAGTGGCCGCGCGCTTCATCTGGGACAGTACTGCACTGCACTGAGCGATG 4260
DB 4201 GAGCAGATGCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4260
QY 140 CCAAGCGCGCAAGCGGATGTGCTTGGCATCTACCAAGCTACAGATGAGACTTTG 199
DB 4261 AGAAGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4320
QY 200 ATCTTGAACCAAGTACGCGGCAAGGCTACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTG 259
DB 4321 TCAACCTGCAAGACTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 4380
QY 260 TTCCCTCGCGGCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 319
DB 4381 TCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4437
QY 320 TGAATGACGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 379
DB 4438 TCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4497
QY 380 TGCTCAGCGGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 439
DB 4498 TACACAGAGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4553
QY 440 ACTTGAAGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 499
DB 4554 --TTTAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4611
QY 500 TCACCATCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 559
DB 4612 TCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4671
QY 560 CGGCGAGGAGCAGATTAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 619
DB 4672 CAGGATGCTCA-----ATAGGCTTGGGACCTGCGCCCTACA 4707
QY 620 TCGCTGAAAGCGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 679
DB 4708 TTCTTGGCCCAATTAATTAAGCTCATGCTGAGGCTGCGCATCTGATCAAGATGATG 4767
QY 680 TTGGCCCTTGCAGAAAGGCGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 729
DB 4768 ACCGCGCAGTCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4827
QY 740 GAGCAGCAATGAGCTTCGCGCAAGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 799
DB 4828 GAGATCCCTCTAACAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 4887
QY 800 GCTGCTTGGCAATCCCATTTTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATG 859
DB 4888 GCTGCTTGGCAATCCCATTTTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATG 4947
QY 860 TGGCGAGAGG 870
DB 4948 TCCGTGACAGG 4958

```

RESULT 5

```
US-10-767-701-6851
; Sequence 6851, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5353)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 6851
; LENGTH: 767
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(767)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS5746_1
US-10-767-701-6851
```

```
Query Match      10.2%; Score 147.8; DB 6; Length 767;
Best Local Similarity 55.4%; Pred. No. 2.5e-30;
Matches 352; Conservative 0; Mismatches 269; Indels 14; Gaps 3;
```

```
QY 16 GCTGCGCCCAAGCATTTGAAATGGGGCTTCGCAAGCGCCGCTTACCAATGGAAGCGCC 75
DB 147 GCGTTCGGAGAGGCGTTCGTTCCGGAAGCGCGGTGCGTACAGGTGAGGGGATG 206
QY 76 GTCAAGAAGATGCGCGCGCGCCGTCATCTGGAGACAGTACTGCGCACTGGAAGCATCG 135
DB 207 GCCAGACAGCGGGGGCGGGGGCCGAGCATCTGGAGCGCTTCATAGAGTACCGGGAGC 266
QY 136 CGCAGCAACGGCGCCACAGCGGATGTGGCTTGCATCTACCTACCGCTTACGATGAGAC 195
DB 267 ATCCCTAACATATGCAACCGCTGATGTGAAGGTGACGATATCACTGGTACAGGAAGAT 326
QY 196 TTTGATCTCTTGACCAAGTACGGCGCAAGGCTTCTCTGTCGTTGTCGCGG 255
DB 327 GTGAACATATGAGAAACATGGCTTTGATGCGTACCGGTTTCAATTTCTTGTCGAGG 386
QY 256 ATCATTTCCCTCGCGCGAGGCTGATCCGTCACAGAGAGGAAATTGATTTTAAAGC 315
DB 387 ATTTTCCCAATGGAATGCGCAAG-----GTGAACAGGAAGAGTGAATTACTATAC 440
QY 316 AACGATGAGAGCGCTTGTAGCGGGGATATCAAGCTTTGGTGACTTTGTACCACTGG 375
DB 441 AGGCTCATATGATTCATGCTTCAGCAAGGTATCAAGCGCGTATGCAAACTCTTACCATTA 500
QY 376 GATGCTCTCAAGGGCTTCAAGATCGCTATGAGAGGCTGCTCAAGTGAAGAGTCCAG 435
DB 501 GACCTCCCATGGAAGCTCAAGTCAAGTATGAGGCTGCTTACG---CCAAAGATTGTG 557
QY 436 CTGACTTTGACGCGATGCAAGGTTGTCTTTTGAACGTTTGGAGACCGAGTCCAGAAC 495
DB 558 GAGGCGTTTGGAGCTACGCGCAAGTTCGCTTCCAGAGCTTGGAGACAGGAGTGAAGAC 617
QY 496 TGAATCACCATCAAGMCCCTGATTCAGGCCATCTATGATATATGCCACCGGAGCAAC 555
DB 618 TGGTTTACCTTCAAGAGCCAGAGTGTGTGCTCTCTTGGCTTACGCAATGGCTTAC 677
QY 556 GCCCGGAGAGAGCATTTAACAGACTCCACGAGGCAACATGCGCACTGAGCCG 615
DB 678 GCANCCGG-----GAAGGTGTCGAGTGGCTCCGAGGCAATCTCAAGAGGAGCCG 732
QY 616 TGGCTCGTGAAGAGGCCAGATCATGAGCCATGC 650
DB 733 TACCTGCTGACACCATCTCATCTTTCATATGC 767
```

```
RESULT 6
PCT-US04-02242-93
; Sequence 93, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING ANI
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; CURRENT FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 93
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-93
```

```
Query Match      9.6%; Score 139; DB 1; Length 1389;
Best Local Similarity 57.1%; Pred. No. 7.2e-28;
Matches 316; Conservative 0; Mismatches 225; Indels 12; Gaps 3;
```

```
QY 11 CGTAGCTCTGCCCAAGCATTTGAATGGGGCTTCGCAAGCGCGCCCTTACCAATGGAAG 70
DB 26 CCTGTCTCTTCCAGAGCAATTCGTCTGGGTCTCTCGCGGCTCTTCAAGTGAAG 85
QY 71 GCGCGCTCAAGAAGTGGCGCGCGCCGTCATCTGGAGACAGTACTGCGCACTGAGAG 130
DB 86 GCGCGCTCAAGAGAGCGGAAGGCCCTTCTGAGACATTTCTGAGAAAGCCG 145
QY 131 CATGCGCACCAACGCGCGCCACAGCGGATGTGGCTTGCATCTACCAACCGCTACAGATG 190
DB 146 GAGCGGCTTCCAGGGGACAGAGGGGCGGTGGCTTGGACCACTATACCGCTACGAG 205
QY 191 AGGACTTGTATCTTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTCTTGTCTGTGT 250
DB 206 AGGACGTGGCTTGTATGCAACAGTGGCGCTGACCGCTTACCGCTGAGGTGTGTGTGC 265
QY 251 CGGGAATCATTTCCCTCGCGCGAGGCTGATCCGTCACAGAGAGGGAATTGATTTT 310
DB 266 CCGAGTGTCTCCGAGAGGGGTGCGGAG-----CCAAAGAGAGGCTTGAATTCT 319
QY 311 ACAGCAAACTGATTTAGCGCCCTTGTAGGCGGGGATATCAAGCTTTGGTACTTTTAC 370
DB 320 ACTCGCGTTGGTGGAGCGGCTGCTGAGGAGGATTTACGCGCTTGGTAACTTTTTC 379
QY 371 ACTGGATCTGCTCAGGGGCTTCAAGATCGCTATGAGAGGCTGAGTCAAGTGAAGAG 430
DB 380 ATTGGAGCTACCCCTTGGCTCT---CTATCAAGGGGGGCTGGCTTCAACCGGAGTAGG 436
QY 431 TCCAGCTGACCTTTGAGCGGTATGCGAGGTTGTCTTTGAAGCTTTTGGGACCGAGTCC 490
DB 437 CGGATTGG---TTTGCAGATACGGGGCTTATAGCGGATGCGCTTCCAGCGGGTGC 493
QY 491 AGAAGTGAATCAACATCAAGMCCCTGATTCAGGCTTATGATATATGCCACCGGCA 550
DB 494 AGCATTTCTTCACTCAAGAGAGCCAGGCTTATATGCGCTTGGACACCTTGAAGGTA 553
QY 551 GCAGCGCCCGGG 563
DB 554 AGCATCTCAAG 566
```

```
RESULT 7
US-10-767-701-14696
; Sequence 14696, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```


Accession	Sequence	Position
Db	GGCAGGAGCGAGAGGGGGTCTATCTCGCTGAGCCTCAGTACACACTGGGAGAGCCCACT	34218
Qy	ACAGCAATGAGCCTCGGGACAAGAGGCTCTGAGCGACGGATGGAATTTCATTTGGCT	802
Db	CACCAAGGGGTCGCCAGAGATGTGAGAGCCGCTACCGAATGCTGCAGTTCTCCCTGGGCT	34218
Qy	GGTTGGCAATGCCCATTTCTTGAGAGAGACTATCCAGAGACATGAAGAGCACTGG	862
Db	GGTTTGTCAACCCCATTTTAGAAGCGAGACTATCTCTACACCAATGAAGTGGAAAGTGG	34338
Qy	GCCGAGAG	870
Db	GGAACAGG	34346

```

: RESULT 11
: US-10-767-701-8210
: Sequence 8210, Application US/10767701
: GENERAL INFORMATION:
: APPLICANT: Kovalic, David K.
: APPLICANT: Zhou, Yihua
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53535)B
: CURRENT APPLICATION NUMBER: US/10/767,701
: CURRENT FILING DATE: 2004-01-29
: NUMBER OF SEQ ID NOS: 63128
: SEQ ID NO 8210
: LENGTH: 689
: TYPE: DNA
: ORGANISM: Sorghum bicolor
: FEATURE:
: OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS33734_1
: US-10-767-701-8210

```

Query Match	8.6%	Score 125.2;	DB 6;	Length 689;
Best Local Similarity	56.4%	Pred. No. 2.9e+24;		
Matches 298;	Conservative	0;	Mismatches 218;	Indels 12;
				Gaps 3;

QY	12	CTAGAGCTGC	CCCAACGACTTGAATG	GGGGGCTTCCGAACGGGCGCCTTACCAAGTCGAAG	71
Db	173	GCAGAGATTTT	CCCAAGGCGCTTGTCTTTGGGGCCGGACATCA	AGCTTATCAGTACGAAG	232
QY	72	CGCGCTCA	AGAGGTGGCCGGCGCCGTCATCTGGGA	CAAGTACTGCACTCGAGCC	131
Db	233	GGCAACCG	ATGATGGAGGAGCCCAAGCAATATGGACA	CGTTTACTCATGCAAGAG	292
QY	132	ATCGCGCA	CCAAAGGCGCCCAAGGGGATGTGGCTTGCCATCACTATCA	TACCACCGCTACGATGA	191
Db	293	GATGCGGA	CAAAACACT--GGGAGTTAGGGGAGAGCGCTAC	CACAAAATACAAAGA	349
QY	192	GGACTTGATCT	CTTGACCAAGTACGGCGCAAGGCGTACCGCTTCTCCTTGC	TCGGTGC	251
Db	350	AGATGTGA	ATTGATGATGATACCTGCGCTCGAAGCGCTACCGCTTTC	CCATTCTTGCTTC	409
QY	252	GCGGATCA	TTCCTCCCTCGCGCGCAGCGCTGATCCCGTCAA	CGAGAGGGAATTGATTTTA	311
Db	410	CAGGCTTCTT	CCAMAGGMAAG-----AGGACCCATCA	CCCCCAAGGGTCTTCATGATTTA	463
QY	312	AGCAAACTGA	TGACGCGCCTGTGTAGGGGGGGGTATCAGCGCTGGGTATCA	TTGTACCA	371
Db	464	CAACAACCTC	ATCAATGACTAATCAAGGGGAATCGAGATACAGTAC	TGACCTGTACCA	523
QY	372	CTGGATCTG	CCTCAGGCGCTTACAGATGCTATGAGGCTGGCTCA	AGTGGMAAGGT	431
Db	524	CTGGATTTCT	CTCAGATCTGTGAAGA	CGATGACCATGCTGGCTCAGCCCCAGGTGGT	583
QY	432	CCAGCTGA	CTTTAGCGGTATGCAAGGTGTGGCTTTGAAACGTTTGGGA	CCGATCCA	491
Db	584	---	GGAGACTTCACGGCGTACGGGACCTGTGCTTCCGGAGTTCTGGG	CGACCGGGTGA	640

Oy 492 GAACGTGATCACCATTCAAGCAWCCCTGGATTTCAGGCCATCTATGATA 539
| | | | | | | | | | | | | | |
Db 641 GCACTGGACGACCATGAGCAGCGCGAAGCTGATCTTCCATCGCGGCCTA 688

```

RESULT 12
US-10-767-795-4264
; Sequence 4264, Application US/10767795
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhao, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 4264
; LENGTH: 649
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: GOSHI-09WAY01-C301_1
US-10-767-795-4264

```

Query Match	8.2%;	Score 119.4;	DB 6;	Length 649;
Best Local Similarity	54.5%;	Pred. No. 1e-22;		
Matches 285;	Conservative 1;	Mismatches 227;	Indels 10;	Gaps 2

Oy	18	CTGTGCCAACA	CACTTTGAAATGGGGCTTCGCAACGGCGCCCTACCAAGATCGAAGGCGCCGT	77
Db	137	TTTTCGGAAGG	GAATTTGTGTTGGAACTGCACACTTCACGTTATCAAGTTGAAGGATGGC	196
Oy	78	CAAGAAGGTG	GCGCGCGCCCGTCATCTGGGACACGATCTGCCACCTCGAAGCCATCGCG	137
Db	197	TAATTAAGAGG	CGCCGAGSACCTTGCACTTTGGGATGTTTATATTAAACAACAGGCCATAT	256
Oy	138	CACCAACGGG	CGCCCAACGGCGATGTGGCTTGCGATCATACTACACCGCTACGATGAGACTT	197
Db	257	TGCTAATATAT	ATATCTGCTGACGTGTGCTGTGACACAGTACACACATTCACAAGGAAGATGT	316
Oy	198	TGATCTCTTG	ACCAAGTACGGCGCAAAAGGCTTACCGCTTCTCTTGTGTGTGGCGGAT	257
Db	317	AGATTTGTGTG	CAAAATTTGCATTTTGATCTTTATCGGTTCTCAATTCATGTCAMGAT	376
Oy	258	CATTCCCTCG	GGGGGACGGCTGATCCCGTCAACGAGGAGGAAATGAGTTTACAGCA	317
Db	377	CTTTCAGAGG	GTG-----TTGAAAGATAAATTTGGAGGAGGTGATTTATACACAG	430
Oy	318	ACTGATTGAC	CCCCCTGTGAGCGGGGATCACGCTTGAGCTTGATTTGTACCACTGGGA	377
Db	431	GTTGATCAAT	TACTTGTGTAAAGAAAGTATTACTCCGATGAGGAACCTTGACATTATAGA	490
Oy	378	TCGTGCTCAG	GCGCTTCACGATCGCTATGAGGCTGCTCAACGTGGAAGGCTCACT	437
Db	491	TCTCCCTCA	AGCTCTTCAGAGAACATGGAATGGGTTCTTGAACCGTCAAA-----TTGTGGA	546
Oy	438	GGACTTTGAG	CGGATGCGAGGTTGTGCTTTGAAACGTTTGGGACCGAGTCCAGACTG	497
Db	547	AAATTAATG	CTGATTATGCAAAAGTTTGTTCMAAAACATTTGGGGGTAAAGTCAAAAACGTG	606
Oy	498	GATCACAAT	CAAGAMCCCTGGAATTCAGGCGCATCTATGGAAT	540
Db	607	GTTTACATT	CAATAAACCAGAAATCGGGGTGTGTTTTGTGTTTT	649

```
RESULT 13
US-10-767-701-12773
; Sequence 12773, Application US/10767701D
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
```


APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 12773
LENGTH: 716
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS25405_1
US-10-767-701-12773

Query Match 7.9%; Score 115.4; DB 6; Length 716;
Best Local Similarity 56.6%; Pred. No. 1.3e-21;
Matches 237; Conservative 0; Mismatches 176; Indels 6; Gaps 1;

QY 20 TGCCCAAGCACTTGAATGGGGCTTCCCAACGCGCCGCTACCAAGATCGAAGCGCCGCTCA 79
DB 281 TCCCAAGGGGTTGTTGTTGGGACGGCGACGTGGCGTACCAAGTGGAGGCGCCCGCT 340
QY 80 AAGAAGGTGGCGCGCCGCTCCATCTGGGACACGTACTGCGCACTGGAGCCATCGCGCA 139
DB 341 CCAACCAAGCGCGGAGACCTCCATCTGGGATTCATTCGCGCACGTCCCAAGAAATATG 400
QY 140 CCAAGCGCGCAAGGGGATGGCTTGCATCTACCTCCACCGGTACAGTAGAGACTTGG 199
DB 401 CGGGAAATCAAAATGGAGACCTTGCATGATCAATCAATGCTACCAAGAAATATG 460
QY 200 ATCTCTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTTGTGCTGGCGATCA 259
DB 461 ATCTCAAGAGATTGATTTGATGCTTACCGGTTTCAATCTGTTGCTCGAGATCT 520
QY 260 TTCCCTCGCGGCGAGCTGATCCCGTCAACGAGAGGGAATTGATTTCACGCAAC 319
DB 521 TCCAGATGGCG-----AGGGAAAGTCAATCCAGAGGATGATATTAACAATATC 574
QY 320 TGATGAGCGCCCTTGGAGGGGATTCACGCTTGGGTGACTTTGATCCACGCGGATC 379
DB 575 TGATTAACATCTCTTCTTCAAGAGCAATGCTCTTACATTAACCTTTACCATATGATC 634
QY 380 TGCCTCAGGCGCTTACGATGCTATGAGGCTGCTCAACGTGAAAGGCTCCAGCTG 438
DB 635 TTCCCTTGGCTTGAAGAAGAAATATGAGGCTTAAAGCGCTAAAGATGGCGGACCTG 693

RESULT 14
US-10-767-701-26695
Sequence 26695, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 26695
LENGTH: 637
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: 30977777
US-10-767-701-26695

Query Match 7.9%; Score 114.8; DB 6; Length 637;
Best Local Similarity 58.4%; Pred. No. 1.7e-21;
Matches 222; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 16 GCTCTGCCCAACGACTTTGAATGGGGCTTGCACACGCGCGCTACCAAGATCGAAGCGCC 75
DB 256 GCGTTCCCAAGAGGGTTTCACTTCGGGACGGCGACGTCCGCTTCCAGGTCCAGAGCGCG 315
QY 76 GTCAAGAGAGTGGCGCGCGCGCTTCATCTGGGACACGTATGCTCCACCTGGAGCCATCG 135
DB 316 GCAACGTCCG 375
QY 136 CGACCAACGCGCGCAACG 195
DB 376 ATGCTGAAACCGGAAACGAGATGTTACCAAGATGATATATCATGCTTACAGGAAGAT 435
QY 196 TTGATCTCTTGAACCAAGTACGCGCGCAAGAGCTTACCGCTTCTTGTGCTGCTCGCG 255
DB 436 GTTATCTCATGAAGAGCTTAATTTGATGATACCGGTTTCAATCTCTGCTGCGAG 495
QY 256 ATCATTTCCCTCG 315
DB 496 ATCTTCCAGATGGCGGAGG-----AAGTTAATGAAGAGATGACGATATTACAC 549
QY 316 AAATGATGACGCGCTTGTGAGCGGGGATACACGCTTGGGTGATCTTGTACCACTG 375
DB 550 AATCTTATGATCAATGATTAAGCAAGTCTTACTCTTAACGCAACCTTAACCACTAT 609
QY 376 GATCTGCTCAGGCGCTTCA 395
DB 610 GATCTTCACTTGGCGCTTCA 629

RESULT 15
US-10-767-795-4474
Sequence 4474, Application US/10767795
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53534)B
CURRENT APPLICATION NUMBER: US/10/767,795
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 117596
SEQ ID NO 4474
LENGTH: 558
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C30142_1
US-10-767-795-4474

Query Match 7.5%; Score 109; DB 6; Length 558;
Best Local Similarity 58.7%; Pred. No. 6e-20;
Matches 225; Conservative 0; Mismatches 155; Indels 3; Gaps 2;

QY 16 GCTCTGCCCAACGACTTTGAATGGGGCTTGCACACGCGCGCTACCAAGATCGAAGCGCC 75
DB 107 GATTTCCCTCCCGCACTTGTCTTGGGTTCATCTTCTTCAAGATGGAAGGAGT 166
QY 76 GTCAAGAGAGTGGCGCGCGCGCGCTTCATCTGGGACACGTATGCTCCACCTGGAGCCATCG 135
DB 167 GTGAACCAAGGTGGCAAGGTAAAGCATATGGATCTTCTTCAATTAAGGAAAA 226
QY 136 CGACCAACGCGCGCAACG 195
DB 227 ATCTTATGGAAGCATGCGATGTTGAGTGAATCATTAACAAGSTACAAAGAGAT 286
QY 196 TTGATCTCTTGAACCAAGTACGCGCGCAAGAGCTTCTTGTGCTGCTGCTGCTGCTGCTG 255
DB 287 ATAGAGCTTATATCCGAGTTAGGTTCAAGCTTCAAGATTTTCCATATCATATGCTCGT 346
QY 256 ATCATTTCCCTCG 315
DB 347 ATTTTCCAGATG--GTTTGGAAAAACAAAGTTATAGGAAGGATTTGCAATTTTACAC 404

Qy	316	AAACTATTGACGCCCCCTGTGAGGGGGGTATCACGCCCTTGGGTGA	CTTTGTACCACTGG	375
Db	405	AATGTATCGATGATGCCCT-TTTGAAAGGGTATGAGCCCTTGTGACATTATACCA	TTGG	463
Qy	376	GATCTGCCCTCAGGCGCTTCACGA		398
Db	464	GATTCTTCCTTGATCTCCATGA		486

Search completed: March 29, 2004, 02:56:08
Job time : 126.98 secs

This Page Blank (uspto)